# FRAGMENT-BASED SCREENING OF THE ONCOGENIC PROTEIN TYROSINE PHOSPHATASE SHP2

by


## LEE DANIEL QUILL


A thesis submitted to the University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

Supervisors: **Professor Michael Overduin, Dr Sam Butterworth and Dr Mark Jeeves**


**The Institute of Cancer and Genomic Sciences**

**College of Medical and Dental Sciences**

**University of Birmingham**


**September 2016**

## Abstract

Protein tyrosine phosphatases represent a family of signalling enzymes with emerging therapeutic potential. The cytoplasmic protein tyrosine phosphatase SHP2, encoded by the *PTPN11* gene, plays a central role in the activation of downstream signalling events at multiple growth factor and cytokine receptors, and was the first oncogenic protein tyrosine phosphatase to be discovered. Aberrant SHP2 signalling underlies the pathology of numerous developmental disorders such as Noonan and LEOPARD syndrome, and is a known driver of breast cancer and myeloproliferative disease. To gain a deeper insight into ligand interactions with the SHP2 catalytic domain in solution, NMR backbone resonance assignments of the 34 KDa SHP2 catalytic domain were determined and utilised in conjunction with $^{15}N$-$^{1}H$ HSQC NMR spectroscopy to map the structurally undisclosed binding site of the previously reported SHP2 inhibitor, NSC-87877. In addition, use of a fragment-based screening approach to accelerate the discovery of novel SHP2 inhibitors has enabled the identification of two novel and distinct chemical scaffolds, both of which now serve as validated chemical precursors for the development of more potent SHP2 lead inhibitors.

## Acknowledgements

Firstly, I would like to thank Professor Michael Overduin for his support, guidance and the opportunity to work on such an exciting and stimulating project. I would also like to issue a huge thanks to Dr Sam Butterworth for agreeing to take over my supervision and providing countless engaging and energising conversations concerning all things drug discovery. I also wish to extend my deepest thanks to Dr Mark Jeeves for his scientific contribution to multiple aspects of my project. Without your guidance, support, and impeccable knowledge of NMR practice and theory, much of the work described in this thesis would not have been possible. Thanks also go to Professor Yotis Senis for the many insightful scientific conversations, entertaining conference trips and the much appreciated career advice and mentorship. Special thanks also go to my AstraZeneca supervisors, Professor Alex Breeze (University of Leeds), Dr Jon Read and Geoffrey Holdgate. Thank you for your supervision, guidance and generously hosting me during my many fleeting visits to complete experiments.

I would also like to take this opportunity to thank the many excellent tutors and lecturers who I have had the privilege of being taught by, both as an undergraduate and doctoral student, during the past seven years here at Birmingham. In particular, I extend my deepest thanks to Dr Klaus Fütterer. Had it not been for your intervention I would not be sitting here writing this thesis. Thank you for nurturing my early enthusiasm for scientific research, for giving me the confidence to rationalise and solve scientific problems independently, and above all, for teaching me to think like a scientist. A special thanks also goes to Dr Eva Hyde for her tireless commitment to teaching me the intricacies of enzyme kinetics and for introducing me to the vast world of NMR spectroscopy through the mind-bending but thoroughly enjoyable tutorials.

Thanks next go to past and present members of the Overduin group for their support, technical assistance and general guidance over the past four years. In particular I thank Dr Timothy Knowles, Dr Marc Lenoir and Dr Claudia Fogl for NMR expertise and technicians Pooja, Sandya and Jas for taking me under their wing and helping me rescue experiments whilst enduring my never-ending questions regarding protein expression and purification.

Much of the research presented in this work would simply not have been possible without the scientific expertise and invaluable contributions from several collaborators, both internally and externally. In particular, I thank Fiyaz Mohammed, Richard Meek and Dr Ian Cadby for crystallography wizardry and the generous allocation of precious synchrotron time for crystal shooting. I also thank Dr Ben Davis (Vernalis) for his scientific expertise regarding fragment-based screening, abundant correspondence to experimental questions and for a providing a valuable industrial perspective on my research.

On a personal note, a huge thank you must go to my family and friends for their continued and unremitting patience and understanding throughout the past 4 years, particularly regarding my general absence from family gatherings and social events. In particular, a special mention goes to my mum and dad for their continuous support and understanding. Words are insufficient to express my gratitude for the sacrifices you have made so that I could be in this position. Thank you for raising me to value the importance of education.

Finally, I extend a heartfelt thank you to Harriet Maydew, my wonderful companion on this journey. Your unwavering love, support and incredible patience sustained this work.

Thank you for all your encouraging and motivating words when I needed them most. I could not have gotten through this period of my life with you.

# Contents

## Abbreviations

| | |
|---|---|
| **AUC** | Analytical Ultracentrifugation |
| **CD** | Circular Dichroism |
| **CML** | Chronic Myeloid Leukaemia |
| **DMSO** | Dimethylsulfoxide |
| **DUSP** | Dual-Specificity Phosphatase |
| **ECL** | Enhanced Luminescence solution |
| **EDTA** | Ethylenediaminetetraacetic acid |
| **FGF** | Fibroblast Growth Factor |
| **Gab2** | GRB2-associated-binding protein 2 |
| **GAP** | GTPase Activating Protein |
| **HEPES** | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| **HEPTP** | Human Haematopoietic PTP |
| **HER2** | Human epidermal growth factor receptor 2 |
| **HRP** | Horseradish Peroxidase |
| **HSQC** | Heteronuclear Single Quantum Coherence |
| **IGF** | Insulin-like Growth Factor |
| **IPTG** | Isopropyl $\beta$-$_D$-1-thiogalactopyranoside |
| **JMML** | Juvenile Myelomonocytic Leukaemia |
| **LB** | Lysogeny Broth |
| **LMPTP** | Low Molecular Weight Protein Tyrosine |
| **LS** | Leopard Syndrome |
| **MAPK** | Mitogen-Activated Protein Kinase |
| **MDS** | Myelodysplastic syndrome |
| **MES** | 2-N-morpholinoethanesulfonic acid |
| **NMR** | Nuclear Magnetic Resonance |
| **NS** | Noonan Syndrome |
| **PDGF** | Platelet-Derived Growth Factor |
| **PIC** | Protease Inhibitor Cocktail |
| **pNPP** | p-nitrophenylphosphate |
| **PRL** | Phosphatase of Regenerating Liver |

| | |
|---|---|
| **PTK** | Protein Tyrosine Kinase |
| **PTP** | Protein Tyrosine Phosphatase |
| **PTEN** | Phosphatase and tensin homologue |
| **PVDF** | Polyvinylidene fluoride |
| **RPTP** | Receptor Tyrosine Phosphatase |
| **SDS PAGE** | Sodium Dodecyl Sulphate Polyacyrlamide Gel Electrophoresis |
| **SH2 domain** | Src Homology 2 domain |
| **SHP1** | Src Homology 2 domain-containing Phosphatase 1 |
| **SHP2** | Src Homology 2 domain-containing Phosphatase 2 |
| **SOC** | Super-Optimal broth |
| **SPR** | Surface Plasmon Resonance |
| **STD NMR** | Saturation Transfer Difference NMR |
| **TBST** | Tris-Buffered Saline and Tween 20 |
| **TCEP** | tris(2-carboxyethyl)phosphine |
| **WaterLOGSY** | Water-Ligand Observed Gradient Spectroscopy |

# List of Figures

# List of Tables

# CHAPTER 1: INTRODUCTION

## 1.1 Phosphorylation as a key cellular signal

Cells have evolved elaborate mechanisms for continuously surveying their internal and external environment, and function by converting these stimuli into a series of chemical messages which programme the appropriate cellular response to a particular stimulus. Reversible protein phosphorylation plays a central role in this process and is a principal mechanism for linking extracellular signal recognition to intracellular signal transduction in higher eukaryotes [1-6]. The balance of protein phosphorylation is controlled by the co-ordinated and reciprocal activities of protein kinases and protein phosphatases which function synergistically to maintain homeostatic levels of cellular phosphorylation. In higher organisms, approximately one third of the total cellular proteome is phosphorylated, mainly on serine, threonine and tyrosine residues [7]. Interestingly, while the dominant contribution to cellular phosphorylation arises through serine and threonine phosphorylation (~98%), tyrosine phosphorylation [8] comprises a mere 2% of the entire phosphoproteome, despite being a fundamental regulatory mechanism underlying a broad spectrum of physiological processes [9-17]. In addition to the critical regulation of normal cellular physiology, disruption of tyrosine phosphorylation levels through perturbing the balance of kinase [18] and phosphatase activity [19] is a hallmark underlying the pathogenesis of numerous human disease states, highlighting the importance of tightly regulated tyrosine phosphorylation in both physiological and pathophysiological contexts [9, 13, 19]. Historically, research centred on protein phosphorylation has placed considerable focus on the study of protein tyrosine kinases (PTKs). PTKs function by catalysing the transfer of a phosphate group onto a particular substrate, and were once considered as the

dominant players in shaping signal responses to external stimuli, with the prevailing notion suggesting that their opposing catalytic partners, protein tyrosine phosphatases (PTPs), functioned merely as simple "housekeeping" enzymes with indiscriminate mechanisms of action and broad substrate specificities [20]. However, the roles of protein tyrosine phosphatases are now known to exhibit far greater functional complexity, and along with their protein kinase counterparts play equally important roles in shaping and regulating cellular signalling responses. Far from being considered merely as passive housekeeping enzymes, it is now acutely apparent that protein tyrosine phosphatases are a highly specific family of enzymes with significant degrees of spatiotemporal regulation [9, 13].

## 1.2 The PTP superfamily

The PTP superfamily is encoded by the largest group of phosphatase genes and are defined by the structurally conserved active site-signature motif $HCX_5R$ ((H)His, (C)Cys, and (R)Arg) in which the cysteine residue functions as a key nucleophile to instigate catalysis. The superfamily of PTP genes can be divided into three discrete protein families based on the sequence similarities of the PTP domains [21]. These include class I PTPs (107 genes), class II PTPs (1 gene), and class III PTPs (3 genes) (Figure 1.1) [9]. The class I PTPs comprise the large majority of human PTP genes and are divided into the classical, phosphotyrosine (pTyr)-specific PTPs and the dual specificity phosphatases (DUSPs). The p-Tyr-specific PTPs (37 genes) are further separated into the transmembrane receptor-like PTPs (RPTPs) and the intracellular non-receptor-like PTPs (NRPTPs) [9, 13, 21]. The DUSP subfamily (~63 genes) are a heterogeneous group of phosphatases exhibiting a greater diversity of substrate specificity and generally have poor sequence similarity beyond the catalytic signature motif. The DUSP phosphatase subfamily share the same catalytic mechanism as the classical PTPs, but the structural architecture of the PTP active site

allows these enzymes to accept phosphoserine (pSer)/phosphothreonine (pThr) as well as pTyr residues [22]. This group also includes the phosphothreonine pThr/pTyr-specific mitogen-activated protein kinase (MAPK) phosphatases, the atypical pThr/p-Tyr- mRNA-specific DUSPs, the phosphoserine (pSer)-specific slingshots, the pTyr-specific phosphatases of regenerating liver (PRLs), the pSer/pThr-specific CDC14s, the PTENs which exhibit specificity for the 3'-phosphate moiety of the inositol ring component of phosphatidylinositol-(3,4,5)-triphosphate, the myotubularin enzymes which catalyse the dephosphorylation of both phosphatidylinositol-3-phosphate and phosphatidylinositol-3,5-bisphosphate, as well as the inositol 4-phosphatases, which mediate hydrolysis of the 4'-phosphate of phosphatidyl inositol-3,4-bisphosphate, inositol-(1,3,4)-trisphosphate, and inositol(3,4)-bisphosphate [9, 21, 22]. Global analysis of the structures of class I PTPs suggest that these enzymes emerged from the evolution of a common ancestor based on the similarities and conserved features of their respective structural folds [20]. The class II family of PTPs appear to conceal a more ancient evolutionary history than their class I PTP family counterparts, and comprise a single gene encoding the pTyr-specific low molecular weight PTP (LMPTP) [23] that contains structural similarities to the overall folds exhibited by the bacterial arsenate reductases. Class II PTP family members are ubiquitously expressed in plants, prokaryotes, and archaea [9]. The class III family of PTPs contain the pThr/pTyr-specific CDC25s, which are suggested to have evolved from bacterial rhodanese-like enzymes (See Table 1.1 for compilation of all known PTP families.) [20].

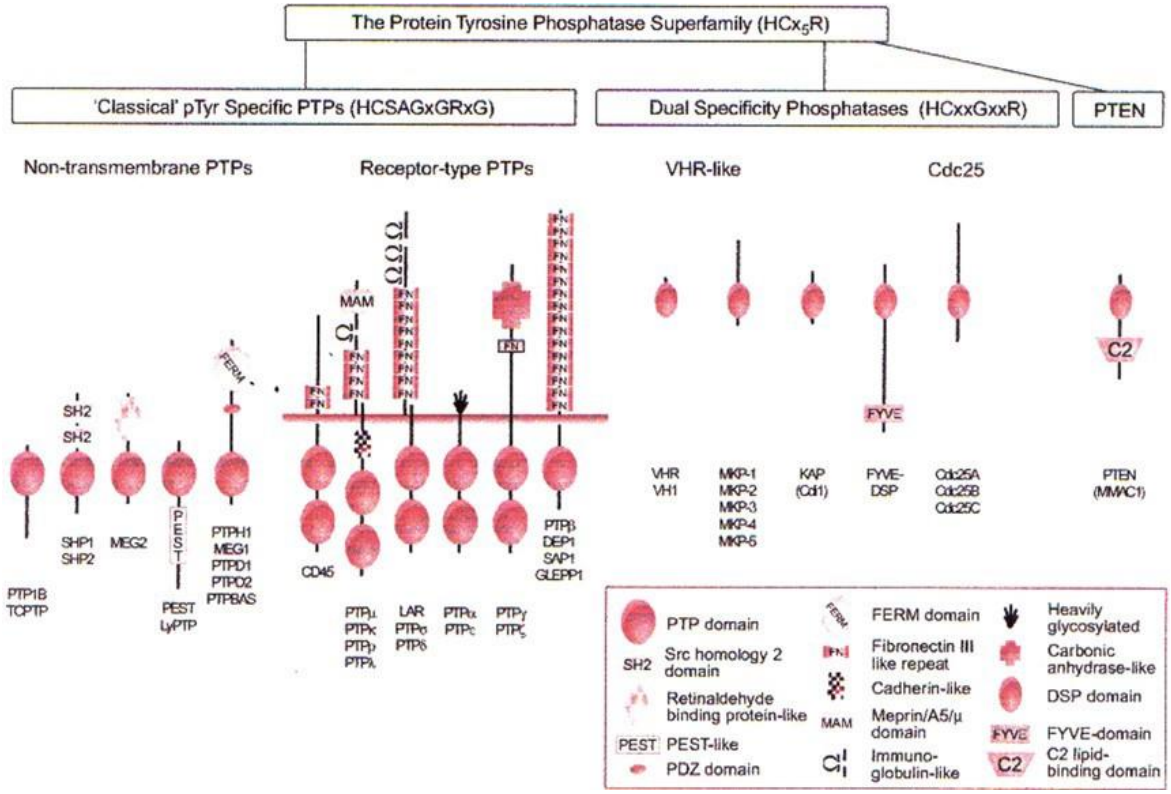



**Figure 1.1 Structural classification of the PTP superfamily**. A schematic representation of the PTP superfamily. The classical phosphotyrosine-specific PTPs (shown on the left) have been further divided into their respective sub-groups: non-transmembrane PTPs and receptor-like PTPs. The right hand side represents the dual-specificity phosphatases (DSPs). PTP catalytic domains are depicted as large and small ovals for classical PTPs and DSPs, respectively. The various signalling modules associated with each phosphatase including regulatory and non-catalytic domains are represented in the box (bottom right) associated with each phosphatase (Figure taken from [24]).

**Table 1.1 Genetic classification of the Human Protein Tyrosine Phosphatome**

| Gene name | Protein, Synonyms | Chromosomal location | Mouse |
|---|---|---|---|
| **A. Class I Cys-based PTPs** | | | |
| **A. 1. Classical PTPs** | | | |
| **A. 1. 1. Transmembrane Classical PTPs** | | | |
| 1. *PTPRA* | RPTPα | 20p13 | YES |
| 2. *PTPRB* | RPTPβ | 12q15-q21 | YES |
| 3. *PTPRC* | CD45, LCA | 1q31-q32 | YES |
| 4. *PTPRD* | RPTPδ | 9p23-p24.3 | YES |
| 5. *PTPRE* | RPTPε | 10q26 | YES |
| 6. PTPRF | LAR | 1p34 | YES |
| 7. *PTPRG* | RPTPγ | 3p21-p14 | YES |
| 8. *PTPRH* | SAP1 | 19q13.4 | YES |
| 9. *PTPRJ* | DEP1, CD148, RPTPη | 11p11.2 | YES |
| 10. *PTPRK* | RPTPκ | 6q22.2-23.1 | YES |
| 11. *PTPRM* | RPTPμ | 18p11.2 | YES |
| 12. *PTPRN* | IA-2, Islet cell antigen | 512 2q35-q36.1 | YES |
| 13. *PTPRN2* | PTPRP, RPTPπ, IA-2β, phogrin | 7q36 | YES<br>YES |
| *14. PTPRO* | GLEPP1/PTP-U2/PTPRO isoforms A/B/C | 12p13.3-p13.2 | YES<br>YES |
| *15. PTPRQ* | PTPS31 | 12q21.31 | YES |
| *16. PTPRR* | PTP-SL, PCPTP, PTPBR7, PC12-PTP1 | 12q15 | YES<br>YES |
| *17. PTPRS* | RPTPσ | 19p13.3 | YES |
| *18. PTPRT* | RPTPρ | 20q12-q13 | YES |
| *19. PTPRU* | PTPJ/PTP-U1/PTPRomicron isoforms 1/2/3 | 1p35.3-p35.1 | YES<br>YES |
| *20. PTPRZ* | RPTPς | 7q31.3 | YES |
| **A. 1.2. Non-receptor PTPs** | | | |
| 21. *PTPN1* | PTP1B | 20q13.1-13.2 | YES |
| 22**.** *PTPN2* | TCPTP, MPTP, PTP-S | 18p11.3-11.2 | YES |
| *23. PTPN3* | PTPH1 | 9q31 | YES |
| *24. PTPN4* | PTP-MEG1, TEP | 2q14.2 | YES |
| *25. PTPN5* | STEP | 11p15.1 | YES |
| *26. PTPN6* | SHP1, PTP1C, SH-PTP1, HCP | 12p12-13 | YES<br>YES |
| *27. PTPN7* | HePTP, LCPTP | 1q32.1 | YES |
| *28. PTPN9* | PTP-MEG2 | 15q23 | YES |
| *29. PTPN11* | SHP2, SH-PTP2, Syp, PTP1D, PTP2C, SH-PTP3 | 12q24.1 | YES<br>YES |
| *30. PTPN12* | PTP-PEST, PTP-P19, PTPG1 | 7q11.23 | YES<br>YES |
| *31. PTPN13* | PTP-BAS, FAP-1, PTP1E, RIP, PTPL1, PTP-BL | 4q21.3 | YES<br>YES |
| *32. PTPN14* | PTP36, PEZ, PTPD2 | 1q32.2 | YES |
| *33. PTPN18* | PTP-HSCF, PTP20, BDP | 2q21.2 | YES |
| *34. PTPN20* | TypPTP | 10q11.22 | YES |
| *35. PTPN21* | PTPD1, PTP2E, PTP-RL10 | 14q31.3 | YES<br>YES |
| *36. PTPN22* | LYP, PEP | 1p13.3-p13.1 | YES |
| *37. PTPN23* | HD-PTP, HDPTP, PTP-TD14, KIAA1471, DKFZP564F0923 | 3p21.3 | YES<br>YES<br>YES |
| **A. 2. DSPs or VH1-like** | | | |
| **A. 2. 1. MKPs** | | | |
| *38. DUSP1* | MKP-1, 3CH134, PTPN10, erp, CL100/ HVH1 | 5q34 | YES<br>YES |
| *39. DUSP2* | PAC-1 | 2q11 | YES |
| *40. DUSP4* | MKP-2, hVH2/TYP1 | 8p12-p11 | YES |
| *41. DUSP5* | hVH3/B23 | 10q25 | YES |
| *42. DUSP6* | PYST1, MKP-3/rVH6 | 12q22-q23 | YES |
| *43. DUSP7* | PYST2, B59, MKP-X | 3p21 | YES |
| *44. DUSP8* | hVH5, M3/6, HB5 | 11p15.5 | YES |
| *45. DUSP9* | MKP-4, Pyst3 | Xq28 | YES |
| *46. DUSP10* | MKP-5 | 1q41 | YES |
| *47. DUSP16* | MKP-7, MKP-M | 12p13 | YES |
| *48. MK-STYX* | MK-STYX | 7q11.23 | YES |
| **A. 2. 2. Atypical DSPs** | | | |
| *49. DUSP3* | VHR, T-DSP11 | 17q21 | YES |
| *50. DUSP11* | PIR1 | 2p13.1 | YE |

**Table 1.1 Continued**

| Gene name | Protein, Synonyms | Chromosomal location | Mouse |
|---|---|---|---|
| 51. *DUSP12* | HYVH1, GKAP<br>LMW-DSP4 | 1q21-q22 | YES |
| 52. *DUSP13A* | BEDP | 10q22.2 | YES |
| 53. *DUSP13B* | TMDP, TS-DSP6 | 10q22.2 | YES |
| *54. DUSP14* | MKP6, MKP-L | 17q12 | YES |
| *55. DUSP15* | VHY, Q9H1R2 | 20q11.21 | YES |
| *56. DUSP18* | DUSP20 LMW-DSP20 | 22q12.2 | YES |
| *57. DUSP19* | *DUSP17*, SKRP1, LDP-2, | 2q32.1 | YES |
| *58. DUSP21* | TS-DSP1 | Xp11.4-p11.23 | YES |
| | LMW-DSP21, BJ-HCC-26 | | YES |
| *59. DUSP22* | tumor antigen | 6p25.3 | YES |
| | VHX, MKPX, JSP1, LMW- | | YES |
| *60. DUSP23* | DSP2, TS-DSP2, JKAP | 11p11.2 | YES |
| | MOSP, similar to RIKEN | | YES |
| *61. DUSP24* | cDNA 2810004N20 | 8p12 | YES |
| *62. DUSP25* | MGC1136 | 1q23.1 | YES |
| *63. DUSP26* | VHZ, FLJ20442, LMW-DSP3 | 2q37.3 | YES |
| | VHP, "similar to RIKEN | | YES |
| *64. DUSP27* | cDNA 0710001B24" | 10q22.3 | YES |
| | DUPD1, FMDSP, "similar to | | YES |
| 65. *EPM2A* | cyclophilin" | 6q24 | YES |
| 66. *RNGTT* | Laforin | 6q16 | YES |
| 67. *STYX* | mRNA capping enzyme | 14 | YES |
| | STYX | | YES |
| **A. 2. 3. Slingshots (3 Genes)** | | | |
| 68. *SSH1* | SSH1, slingshot 1 | 12q24.12 | YES |
| *69. SSH2* | SSH2, slingshot 2 | 17q11.2 | YES |
| *70. SSH3* | SSH2, slingshot 3 | 11q13.1 | YES |
| **A. 2. 4. PRLs (3 Genes)** | | | |
| 71. *PTP4A1* | PRL-1 | 6q12 | YES |
| *72. PTP4A2* | PRL-2, OV-1 | 1p35 | YES |
| *73. PTP4A3* | PRL-3 | 8q24.3 | YES |
| **A. 2. 5. CDC14s (4 Genes)** | | | |
| 74. *CDC14A* | CDC14A | 1p21 | YES |
| *75. CDC14B* | CDC14B | 9q22.33 | YES |
| *76. CDKN3* | KAP | 14q22 | YES |
| *77. PTP9Q22* | PTP9Q22 | 9q22.32 | YES |
| **A. 2. 6. PTENs (5 Genes)** | | | |
| 78. *PTEN* | PTEN, MMAC1, TEP1 | 10q23.3 | YES |
| 79. *TPIP* | TPIPα, TPTE and PTEN | 13q12.11 | YES |
| | Homologous | | YES |
| 80. *TPTE* | PTEN-like, PTEN2 | 21p11 | YES |
| 81. *TNS* | Tensin | 2q35-q36 | YES |
| 82. *TENC1* | C1-TEN, TENC1, | 12q13.13 | YES |
| | KIAA1075 | | |
| **A. 2. 7. Myotubularins (15 Genes)** | | | |
| 83. *MTM1* | myotubularin | Xq28 | YES |
| 84. *MTMR1* | MTMR1 | Xq28 | YES |
| 85. *MTMR2* | MTMR2 | 11q22 | YES |
| 86. *MTMR3* | MTMR3, FYVE-DSP1 | 22q12.2 | YES |
| 87. *MTMR4* | MTMR4, FYVE-DSP2 | 17q22-q23 | YES |
| 88. *MTMR5* | MTMR5, SBF1 | 22q13.33 | YES |
| 89. *MTMR6* | MTMR6 | 13q12 | YES |
| 90. *MTMR7* | MTMR7 | 8p22 | YES |
| 91. *MTMR8* | MTMR8 | Xq11.2 | NO |
| 92. *MTMR9* | MTMR9, LIP-STYX | 8p23-p22 | YES |
| 93. *MTMR10* | MTMR10 | 15q13.1 | NO[d] |
| *94. MTMR11* | MTMR11CRA α/β | 1q12.3 | YES |
| *95. MTMR12* | MTMR12, 3-PAP | 5p13.3 | YES |
| *96. MTMR13* | MTMR13, SBF2, CMT4B2 | 11p15.3 | YES |
| *97. MTMR14* | FLJ22075, hJumpy, | 3p26 | YES |
| | hEDTP | | |

| Table 1. Continued | | | | |
|---|---|---|---|---|
| **Gene name** | **Protein, Synonyms** | **Chromosomal location** | | **Mouse** |
| **B. Class II Cys-Based PTPs (1 Gene)** | | | | |
| 98. *ACP1* | LMPTP, low Mr PTP, LMWPTP, BHPTP | 2p25 | | YES |
| **C. Class III Cys-Based PTPs (3 Genes)** | | | | |
| 99. *CDC25A* | CDC25A | 3p21 | | YES |
| 100. *CDC25B* | CDC25B | 20p13 | | YES |
| 101. *CDC25C* | CDC25C | 5q31 | | YES |
| **D. Asp-Based PTPs (4 genes)** | | | | |
| 102. *EYA1* | Eya1 | 8q13.3 | | YES |
| 103. *EYA1* | Eya2 | 20q13.1 | | YES |
| 104. *EYA1* | Eya3 | 1p36 | | YES |

**Table 1.1. A compilation of all known PTPs currently cited in the scientific literature**. Current data indicate that there are 104 genes encoding PTPs in humans, including 100 class I, 1 class II, and 3 class III PTP genes. The earlier predicted human OST-PTP has been omitted here. Although previously reported as a functionally active PTP in mouse and rat, the human OST-PTP cDNA sequence contained numerous disablements and upon closer examination was classified as a pseudogene [25]. The earlier predicted myotubularin-related phosphatase MTMR15 [10] has also been omitted here. This protein (KIAA1018/FAN1) was functionally assigned as a DNA repair nuclease [26].

## 1.3 The PTP catalytic mechanism

The universally conserved mechanism of catalysis exhibited by all PTPs was first elucidated by a combination of site-directed mutagenesis, kinetic analysis [27-29] and high resolution structural information enabling complementary insights into the biochemical machinery driving PTP catalysis [30-34]. This uncovered the structural and biochemical basis underpinning PTP activity, and revealed that PTP family members share a highly conserved catalytic mechanism centred on a nucleophilic cysteine residue acting as the initial mediator of phospho-substrate hydrolysis. The highly unstable transition state of the PTP reaction is stabilised by a key arginine residue, with hydrolysis of the phospho-intermediate mediated by a catalytic acid-base aspartate. The overall reaction is orchestrated by a number of conserved residues optimally positioned in a series of conserved loop regions which together form a pre-organised and catalytically competent active site primed for the acceptance and turnover of phospho-containing substrates. The first step of phospho-substrate hydrolysis features the initial cleavage of phosphorous-oxygen bond via a nucleophilic attack on the phosphorous atom by the catalytic cysteine residue in the phosphate-binding loop (P-loop). During this process, the catalytic aspartate residue present in the WPD-loop functions as a general acid by readily donating a proton to stabilise the $OH^-$ leaving group. This step generates a phospho-cysteine intermediate along with the concomitant release of the dephosphorylated substrate. The second rate limiting step of the reaction features the hydrolysis of the thiophosphate intermediate assisted by the catalytic aspartate in the WPD-loop which acts as the general base by facilitating deprotonation of a key water molecule. Abstraction of the water proton creates the hydroxyl nucleophile which cleaves the phospho-sulphur bond to regenerate the catalytic cysteine, release free phosphate and perpetuate the catalytic cycle. In classical

PTPs, the nucleophilic water molecule is co-ordinated and positioned by a conserved glutamine located in the Q-loop. An additionally important feature concerns the guanidinium side chain of a conserved arginine residue located in the P-loop which plays a crucial role in both stages of the phosphatase catalytic mechanism. Firstly, the guanidinium side chain acts by stabilising the transition state of the reaction by co-ordinating the phosphate group during substrate recognition and nucleophilic attack, and also stabilises the phospho-cysteine intermediate during the second nucleophilic attack which liberates the free phosphate (Figure 1.2). Although essential for stabilisation of the transition state, mutational studies have revealed that PTPs lacking both aspartate and arginine retain catalytic activity, but mutation of the critical cysteine results in complete abrogation of phosphatase activity [35]. Despite exhibiting a highly conserved active site, different PTPs in the same subfamily demonstrate different catalytic turnover rates toward the generic phosphatase substrate *p*-nitrophenyl phosphate. Interestingly, the Yersinia PTP YopH [36] is approximately 10-fold more active than human hematopoietic PTP (HePTP) [37]. Thus, despite the highly conserved active site architecture and chemistry, modest shifts in the active site microenvironment can significantly perturb the efficiency of PTP catalysis. The efficiency of PTP catalysis is also greatly influenced by the conformational flexibility of the WPD-loop, which transitions between an open conformation in the apo-form of the enzyme, and a closed conformation upon substrate and ligand binding to the active site [38]. Indeed, both targeted mutations and ligands which reduce WPD-loop conformational dynamics result in compromised PTP catalytic activity [39, 40].

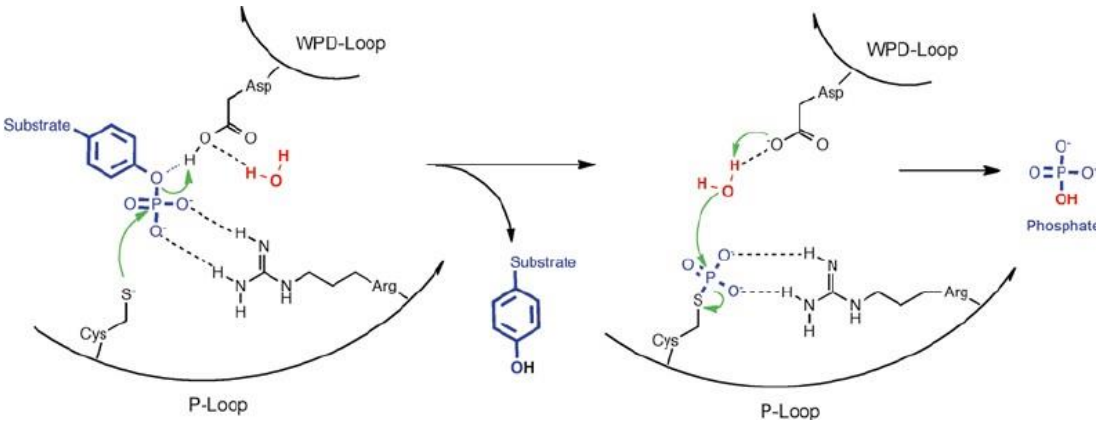



**Figure 1.2 The conserved catalytic mechanism of protein tyrosine phosphatases.** The nucleophilic cysteine residue present in the P-loop instigates the cleavage of the phosphorous-oxygen bond in the phosphorylated substrate. At the same time, the catalytic aspartate residue located in the WPD loop acts as a general acid catalyst by donating a proton to the hydroxyl leaving group. This step generates a phosphocysteine intermediate and the unphosphorylated substrate is ejected from the PTP active site. The phosphorous sulphur bond then undergoes hydrolysis via a hydroxyl ion which was generated with the assistance of the WPD-loop catalytic aspartate which abstracts the proton from a water molecule to create the hydroxyl nucleophile. Figure taken from [19].

## 1.4 Structural architecture of the PTP domain

Over the past two decades significant progress has been made in unravelling the functional importance of phosphatase activity in signal transduction. A large part of this effort has focussed on dissecting the structural architecture of the phosphatase active site and mapping the key determinants of PTP specificity at the molecular level. Multiple atomic resolution structures of human PTP catalytic domains have now been solved spanning several different families and subfamilies. The archetypal PTP, PTP1B, was the first structure solved at high resolution and provided the first atomic-level insights into the structural architecture of the PTP catalytic pocket [31]. The overall fold of classical PTP domains is extremely well conserved and comprises of approximately 280 amino acid residues interspersed with a series of defined short sequence motifs (Figure 1.3). These short sequence motifs are characterised by several loop regions which collectively form the canonical active site architecture of all classical PTPs. Such loop regions include: (1) the phosphate-binding loop (P-loop), which contains the PTP signature motif harbouring the catalytically critical cysteine and arginine residues; (2) the WPD-loop which contains the catalytic acid/base aspartate; (3) the Q-loop which harbours the conserved glutamine residue for co-ordination of the water molecule prior to nucleophilic attack of the phospho-cysteine and ejection of free phosphate from the PTP active site; (4) the  for pTyr in classical PTPs; and 5) the E-loop, which harbours several conserved residues that play a key role in controlling the dynamics of the catalytically critical WPD-loop (Figure 1.3). Interestingly, the pTyr recognition loop is only found in classical pTyr-specific PTPs, where it functions by defining the deep catalytic cleft which enables the essential discrimination between pTyr and pSer/pThr residues [41].

### 1.4.1 The P-loop

The base of the catalytic cleft present in all classical PTPs is formed by the P-loop, which harbours the essential PTP signature motif $HC(X_5)R$. The P-loop is structurally and chemically optimised for accepting pTyr substrates, with backbone amide nitrogen atoms and the guanidinium side chain of the conserved P-loop arginine residue providing a complementary positively charged microenvironment. The evolution of the catalytic pocket toward favouring a positively charged microenvironment is critical for maximising the affinity for the phosphate anion, and also serves to maintain the nucleophilicity of the catalytic cysteine by stabilising the thiolate anion ($Cys\text{-}S^-$). While the sulphydryl side chains of cysteine residues are not typically nucleophilic under physiological conditions (with pKa values of ~8.5), the positively charged microenvironment of the phosphatase active site creates an unusually low pKa value (typically between 4.5-5.5) allowing the reaction to proceed under physiological conditions [42]. However, the low pKa renders the catalytic cysteine exquisitely sensitive to oxidation [43], nitrosylation [44], and sulfhydration [45] which can lead to both reversible and irreversible modifications that perturb nucleophilicity and lead to overall cessation of enzyme activity.

### 1.4.2 The WPD-loop

The WPD-loop is a structural motif present in many PTPs (the exception being the myotubularins and the CDC25 family), and is located approximately 30-40 residues upstream of the PTP signature catalytic motif in pTyr-specific PTPs. Both the tryptophan and proline residues are 100 and 97% conserved across the classical pTyr-specific PTPs, and play critical roles in governing WPD-loop dynamics and facilitating PTP catalysis. In classical PTPs, the tryptophan residue in the WPD loop has been identified as an essential

hinge residue and thus appears to be an essential determinant of WPD-loop flexibility [46, 47]. In contrast, the aspartate residue functions as the general acid/base during PTP catalysis. The central role of the WPD-loop in PTP catalysis is to act as a flexible gate which regulates substrate access to the active site. As such, the WPD-loop has been observed in both closed (active) conformations as well as open (inactive) conformations [35]. In the absence of pTyr-containing substrates, the WPD-loop continuously cycles between active and inactive conformations [48]. However, substrate binding is only possible when the loop adopts an open conformation. Upon recognition of pTyr-containing substrates, the WPD-loop transitions to the closed conformation which positions the catalytic aspartate residue in close proximity to the catalytic cysteine in order to facilitate catalysis. Indeed, mutational studies investigating targeted mutations of the tryptophan hinge residue in the WPD-loop result in impaired conformational mobility which leads to compromised PTP catalytic activity. Interestingly, PTPs with a different amino acid substituted in position of the aspartate, including the receptor protein tyrosine phosphatase (RPTP) IA-2 (PTPRN; D->A) [49] exhibit very low activity levels, and in some cases are inactive.

### 1.4.3 The p-Tyr recognition loop

The p-Tyr-loop, often referred to as the substrate-binding loop, is a conserved structural feature of all classical PTPs, and plays a central role in defining the depth of the phosphatase catalytic pocket. The structural elements forming the architecture of p-Tyr loop determine the selectivity and specificity necessary for optimal recognition of p-Tyr-containing substrates, and provide the mechanistic basis for discriminating p-Tyr and pSer/pThr-containing substrates. These structural elements include the loop region-

containing p-Tyr-recognition KNRY motif, which features a highly conserved tyrosine residue optimally positioned to provide the appropriate 9Å depth critical for pTyr recognition. Aside from its role in priming the catalytic pocket for optimal p-Tyr specificity, the conserved tyrosine residue plays a secondary role in forming pi-pi stacking interactions with the p-Tyr moiety on substrate engagement, thus consolidating substrate interaction in the active site [31]. The conserved arginine residue in the KNRY motif appears to stabilise the overall loop dynamics and serves to anchor the substrate through side-chain interactions between the guanidinium group and nearby backbone oxygen atoms. Interestingly, despite the essential nature of the tyrosine residue for KNRY motif function and overall substrate specificity, a total of six phosphatases including PTPRN, PTPRN2, PTPRQ, PTPRU and PTPN21 have replaced tyrosine for an alternative amino acid at this position [21]. Furthermore, the phosphatases for which these substitutions are apparent also exhibit alterations in their WPD motifs, which may reflect a change in the functional trajectory of these enzymes throughout the course of evolution. For conventional KNRY motif-containing phosphatases, additional structurally-conserved features include aspartate or asparagine residues positioned two-residues from the C-terminal end of the conserved tyrosine. The functional importance of this residue hinges on the formation of a dual hydrogen bond interaction between the backbone amide nitrogen atoms of the pTyr substrate and adjacent residue providing additional stabilisation of the p-Tyr substrate interaction.

### 1.4.4 The E-loop

The functional significance of the E-loop is defined by the presence of key glutamate and lysine residues which are universally conserved across all classical PTPs. In approximately 85% of all determined structures of classical PTPs, the E-loop is found as a tight β-hairpin motif with the critical glutamate residue forming a dual hydrogen bonding interaction with the guanidinium side chain of the invariant arginine residue present in the P-loop. This interaction is thought to provide sufficient stabilisation of the guanidinium group, thus creating a geometrically favourable orientation for phosphate binding to the P-loop. Although the E-loop exhibits a spectrum of conformational flexibility [50-52], and in some cases appears intrinsically disordered [53] the glutamate-arginine interaction remains a highly conserved feature found in almost all reported crystal structures of classical PTPs. In addition to the functional role exhibited by the conserved glutamate residue, the conserved E-loop containing lysine residue forms a hydrogen bonding interaction with the closed-conformation of the WPD-loop, specifically through the catalytically critical aspartate residue, which likely serves to stabilise the WPD in its substrate bound conformation. Mutational studies probing the functional significance of the lysine residue confirmed that creating the alanine mutant reduces the catalytic activity of HePTP [54]. Despite the key functional contribution of the E-loop in the overall mechanism of phosphate binding, observations of similar functions outside the classical PTP family have been surprisingly elusive. Although the E-loop is similarly positioned across all PTP subfamilies, the conserved positions and mechanisms of action of glutamate and lysine residues are only found in pTyr-specific PTPs.

### 1.4.5 The Q-loop

The Q-loop is a conserved structural element underpinning all classical PTPs, and comprises a highly conserved glutamine residue which plays a fundamental role in the PTP catalytic mechanism. In particular, a combination of kinetic [55-57] and structural analyses [27-29, 31, 44] of PTPs revealed that the glutamine side chain extends into the central cavity of the catalytic pocket, serving to co-ordinate and position the key nucleophilic water molecule prior to phospho-thioester hydrolysis. Substitution of the conserved glutamine with a residue devoid of hydrogen bonding capacity with the nucleophilic water molecule results in phosphotransferase activity as opposed to strict hydrolysis of phosphoester substrates exhibited by classical PTPs. Therefore, the Q-loop imparts the strict and specific hydrolytic activity manifested by classical PTPs, preserving PTP activity and thus preventing the acquisition of phosphotransferase activity exhibited by kinases. In addition to maintaining the strict hydrolytic activity reminiscent of classical PTPs, the nucleophilic water molecule (positioned by the conserved Q-loop glutamine residue) has also been proposed as a key player in WPD-loop closure prior to nucleophilic attack of the pTyr-containing substrate [33].

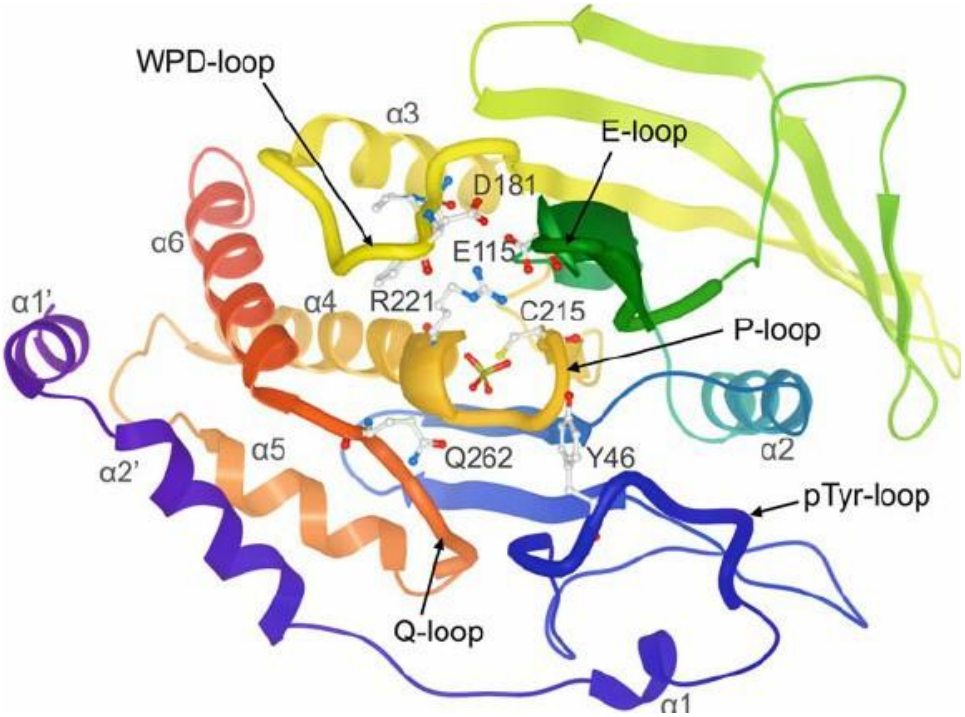



**Figure 1.3 Ribbon diagram of the classical architecture exhibited by a class I PTP catalytic domain**. Structural representation of a class I PTP catalytic domain with a tungstate ion bound in the active site (PTP1B; PDB: 2HNQ). Conserved amino acid residues are highlighted as ball-and-stick representation and are identified in accordance with their location within key structural loops. These residues include: the catalytic cysteine (C215) and conserved arginine (R221) of the P-loop; the acid-base aspartate residue (D181) of the WPD-loop; a conserved glutamine (Q262) residing in the Q-loop; an essential tyrosine (Y46) present in the pTyr-recognition loop, and a conserved glutamate (E115) of the E-loop. Figure taken from [41].

## 1.5 The Protein Tyrosine Phosphatase SHP2

### 1.5.1 Discovery of SHP2

The protein tyrosine phosphatase SHP2 was initially identified by several groups as a Src-homology 2 (SH2)-domain containing protein tyrosine phosphatase [58-62]. Following its biochemical classification, further genotypic analysis revealed that the Src homology 2 domain-containing protein tyrosine phosphatase (SHP2) belonged to a relatively small sub-family of non-receptor protein-tyrosine phosphatases which comprised two human members, Src homology 1 domain-containing protein tyrosine phosphatase (SHP1; encoded by the *PTPN6* gene) and SHP2 (encoded by the *PTPN11* gene). Interestingly, global analysis of the prevalence of SHP genes across higher organisms revealed that all vertebrate organisms have ortholog copies for both SHP genes, whereas invertebrates such as *Drosophilia* and *Caenorhabditis elegans* contain single variants of the SHP gene known as *csw* and *ptp-2*, respectively. Both *csw* and *ptp-2* exhibit SHP2-like functions suggesting that SHP1 may have undergone evolutionary divergence to facilitate a more specialised functional role. This is supported by the restricted expression pattern of SHP1 to haematopoietic cells, whereas SHP2 displays a more ubiquitous expression pattern across multiple tissues [63, 64].

The disease-linked propensities of SHP2 were first highlighted by observations of germline mutations in the PTPN11 gene. These genetic aberrations were first identified in Noonan Syndrome (NS) [65], a common autosomal dominant developmental condition characterised by a spectrum of phenotypic alterations, including facial abnormalities, defective cardiac function, retarded growth patterns and severe learning difficulties [66-69]. NS has been estimated to affect around 1 in 1000-2000 patients, with germline PTPN11

mutations reported in 40-50% of NS patients [70]. Similar germline PTPN11 mutations also underlie related but much rarer genetic disorders, including Noonan-like disorder with multiple giant lesion syndrome and LEOPARD syndrome [66]. In addition to germline mutations, somatic mutations in PTPN11 have been identified in ~35% of cases of juvenile myelomonocytic leukaemia (JMML) [71, 72], an aggressive childhood neoplasm defined by excessive and deregulated production of monocytic cells that drive invasion of the GI tract, lung, liver and spleen. Further somatic mutations have also been found in solid tumours [70], as well as other types of myeloproliferative neoplasm [73-75].

Collectively, the prevalence of both germline and somatic mutations across a broad range of SHP2-linked pathologies culminated in the identification of SHP2 as the first proto-oncogenic PTP [71].

### 1.5.2 Molecular structure and regulation of SHP2

The domain organisation and overall structure of SHP2 is composed of two SH2 domains (N-SH2 and C-SH2), a globular PTP domain, a disordered C-terminal tail containing tyrosine phosphorylation sites (Y542 and Y580), and a proline-rich motif [61] (Figure 1.4) While the majority of the PTP family are generally considered as negative regulators of signalling pathways, due to their ability to reverse phosphorylation by PTKs and terminate the lifetime of a signal, a large body of genetic and biochemical evidence suggests that SHP2 plays a central role in the propagation and activation of signalling cascades at a diverse range of receptor tyrosine kinases (RTKs) [76] and cytokine receptors [71], and is thus required for full activation of the RAS/MAPK pathway [24, 77]. Furthermore, almost all signalling capabilities and biochemical functions of SHP2 are acutely dependant on the presence of functional and intact SH2 and PTP domains [14, 61]. Despite extensive

characterisation of SH2 and PTP domain structure and function in the overall context of SHP2 signalling, the structural significance of the disordered C-terminus of SHP2 remains comparatively underexplored. However, at a functional level, the phosphorylation state of the C-terminal tyrosine residues has been shown to arise in response to receptor-specific SHP2-signalling. In particular, C-terminal tyrosine residues are transiently phosphorylated upon activation of fibroblast growth factor (FGF) and platelet-derived growth factor (PDGF) receptor signalling, but exist in the unphosphorylated form in response to epidermal growth factor (EGF) and insulin-like growth factor (IGF) receptor signalling, suggesting an alternative modulatory mechanism in EGF and IGF-mediated signalling pathways [78]. Although relatively uncharacterised, the proline-rich motif has been shown to bind the SH3-domain of Src kinase enabling transient increases in Src kinase activity [79]

Under basal conditions, in the absence of mitogenic stimuli, SHP2 adopts an autoinhibited conformation mediated by an intramolecular interaction between the N-SH2 domain and the PTP domain [80, 81]. Upon mitogenic stimulation by growth factors or cytokines, the autoinhibition of SHP2 is relieved through interaction of the tandem SH2 domains (Figure 1.5) with pTyr residues in the cytoplasmic domains of RTKs or scaffolding adaptors, such as insulin receptor substrates, fibroblast growth factor receptor substrate, or GRB2-associated binders (GAB) proteins. Engagement of pTyr instigates a conformational change which dislodges the occluding N-SH2 domain and facilitates substrate access to the exposed catalytic site [78, 79]. The molecular basis underlying this auto-inhibitory mechanism of SHP2 was first illuminated by structural studies revealing the high resolution crystal structure of full-length SHP2. The crystal structure of full length SHP2 revealed an overall compact structure with the SH2 domains organised around the PTP domain with the phosphopeptide binding sites exposed in the correct orientation for

substrate recognition [78]. This intricate catalytic mechanism of SHP2 can thus be thought of as a molecular switch which, under appropriate extracellular cues and through localisation to the correct signalling receptors, can efficiently couple catalytic activation to the propagation of a specific physiological stimulus [72]. The physiological significance of the intrinsic regulatory mechanism of SHP2 was confirmed by mutational studies targeting the residues lining the SH2-PTP interface. Engineering these mutations to disrupt the interface and destabilise the SH2-PTP interaction generated mutants with enhanced and deregulated catalytic activity [82].

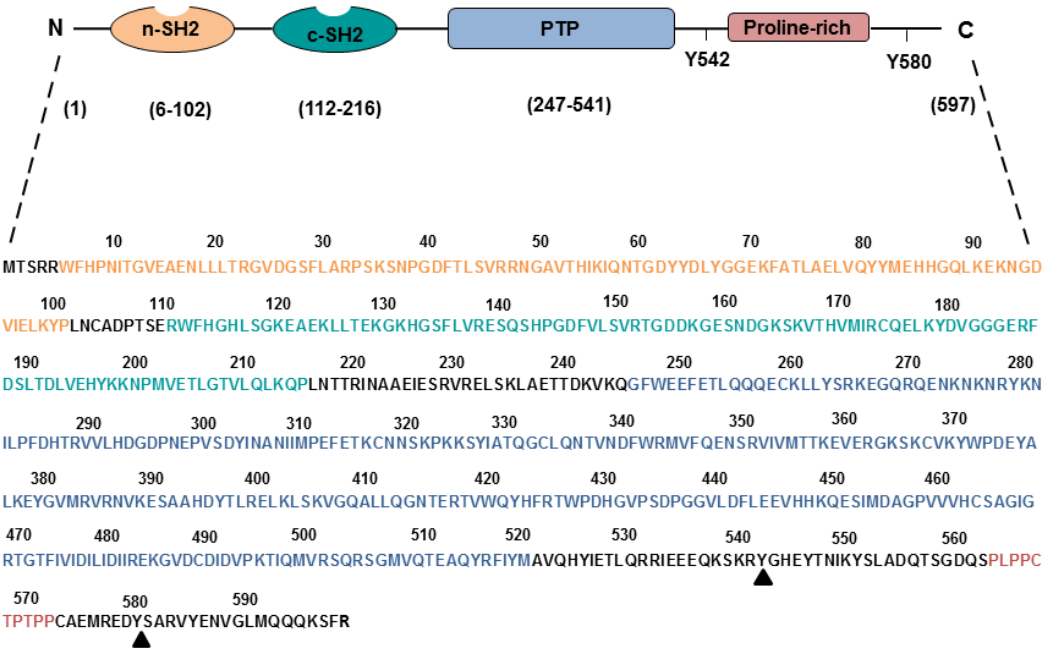



**Figure 1.4 Domain organisation of full length SHP2.** Schematic representation of the modular domain architecture of full length SHP2 (amino acids 1-597). The structure of Shp2 consists of two N-terminal SH2 domains designated as n-SH2 (highlighted orange; amino acids 6-102) and c-SH2 (highlighted dark teal; amino acids 112-216), a globular catalytic (PTP) domain (highlighted blue; amino acids 247-541) and a largely unstructured C-terminal tail harbouring a proline-rich region (highlighted red) and two phosphorylation sites (Y542 and Y580). The amino acid sequence of full length SHP2 has been coloured to highlight individual domains. Black triangles indicate phosphorylation sites present within the C-terminal tail.

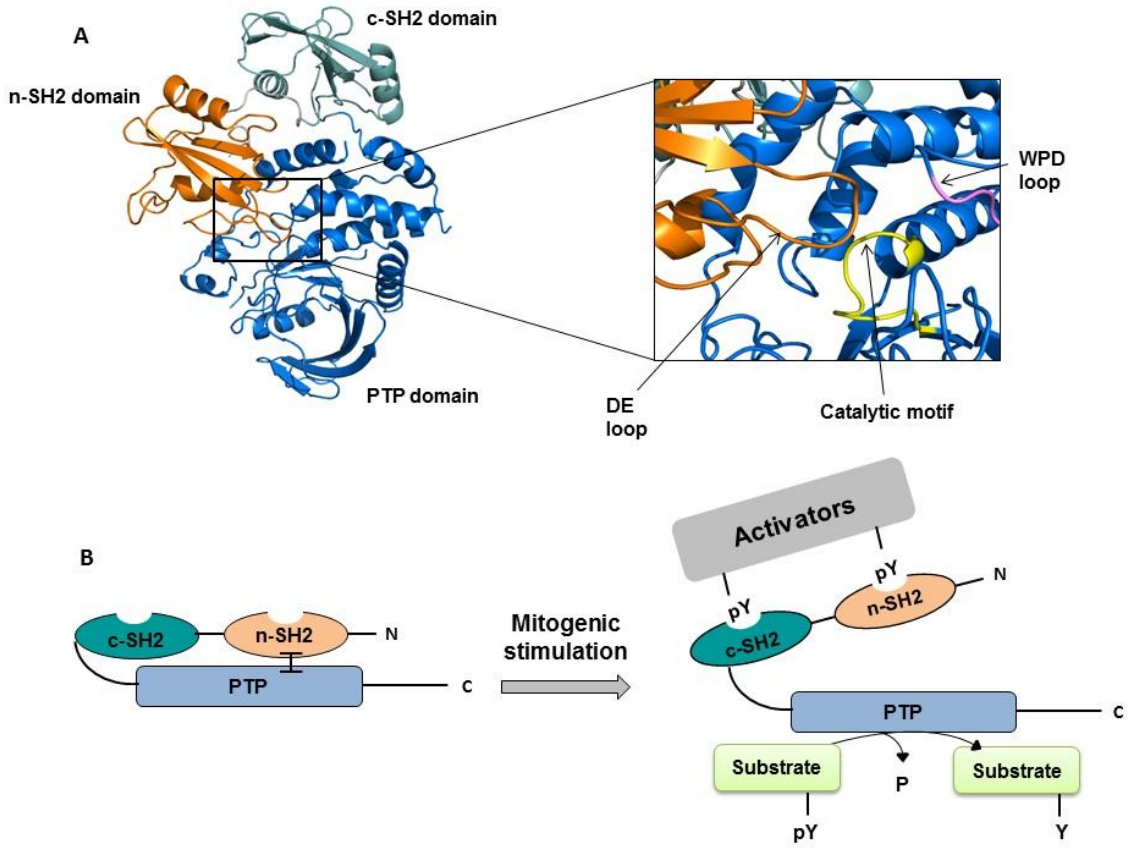



**Figure 1.5 Mechanism of auto-inhibition of SHP2.** A. Molecular structure of SHP2 in the auto-inhibited conformation (PDB: 2SHP) [80]. The structural basis for SHP2 auto-inhibition was first identified from a crystal structure of full length SHP2 with a truncated C-terminal tail. This revealed the primary role of the N-SH2 which mediates auto-inhibition of SHP2 activity via conformational occlusion of the PTP catalytic site. In particular, the DE loop of the N-SH2 domain projects into the PTP catalytic site and forms multiple contacts with residues present in the catalytic motif and WPD loop which help to stabilise the auto-inhibited conformation. B. The release of SHP2 auto-inhibition is mediated by a diverse cascade of physiological stimuli. At basal levels of cellular stimulation, SHP2 exists primarily in the auto-inhibited conformation. Upon mitogenic cellular stimulation, the auto-inhibition of SHP2 is relieved through engagement of the N-SH2 and C-SH2 domains and subsequent dephosphorylation of phosphorylated tyrosine motifs present the cytoplasmic tails of activated growth-factor receptors. Disengagement of the N-SH2 domain through phosphotyrosine binding exposes the SHP2 catalytic site and promotes SHP2 activation and downstream propagation of growth factor signalling pathways.

### 1.5.3 SHP2 signalling and substrates

Although catalytically active SHP2 is required to initiate downstream activation of the RAS/MAPK pathway (Figure 1.6), identification of the precise mechanism and key signalling targets driving activation of Ras signalling has, until recently, remained unclear, with various pathways proposed [83, 84]. One proposed signalling mechanism for SHP2-mediated dephosphorylation was the pTyr sites on RTKs involved in binding either p120 RasGAP or the phosphorylation sites on the p120 RasGAP binding protein, Gab1 [85-88]. In this instance, SHP2-mediated dephosphorylation would abrogate GTPase activity and thus act to promote Ras activation. An additional signalling mechanism postulated for triggering RAS/MAPK signalling concerns the SHP2-mediated activation of Src family kinases through inhibition of C-terminal Src kinase [89]. This indirect mechanism is suggested to occur via the dephosphorylation of the C-terminal Src kinase binding protein or paxillin. Other signalling targets postulated as sites for SHP2 dephosphorylation include RASGAP binding sites on a selection of RTKs [90, 91] and the tyrosine phosphorylation sites on SPROUTY proteins [92, 93]. In addition, SHP2 may also facilitate Ras activation through the regulation of transcription factors responsible for downregulation of the Ras GTPase activating protein, NF1, which acts as a negative regulator of Ras signalling [60].

More recently, however, an alternative mechanism of Ras/MAPK pathway regulation has been demonstrated through Src kinase-mediated phosphorylation on the conserved Y32 residue located in the switch I region of the Ras GTPase. Furthermore, Src-mediated phosphorylation at Y32 promotes uncoupling of the Ras-Raf interaction leading to enhanced GAP recruitment, increased GTPase activity and inactivation of Ras [94]. Interestingly, SHP2 has been identified as a key mediator of direct Ras activation through

dephosphorylation of Y32, thereby reversing the negative regulation of Ras activity imposed by Src-phosphorylation and promoting continuation of the Ras-GTPase cycle [95].

In addition to its established and emerging roles in Ras/MAPK pathway regulation, SHP2 also appears to exert cell-type specific effects on a myriad of other signalling pathways. In particular, SHP2 has been implicated in negative regulation of the PI3K/AKT pathway in response to EGFR stimulation [96, 97], and also plays central roles in PDGF and IGF-mediated AKT activation [94]. Furthermore, observation of the signalling activities of SHP2 also extend to the regulation of JNK, NF-kB [98], Rho [99, 100] and NFAT [101] activation, as well as additional studies implicating contrasting inhibitory and activatory roles for wild type SHP2 in JAK/STAT signalling [61, 75].

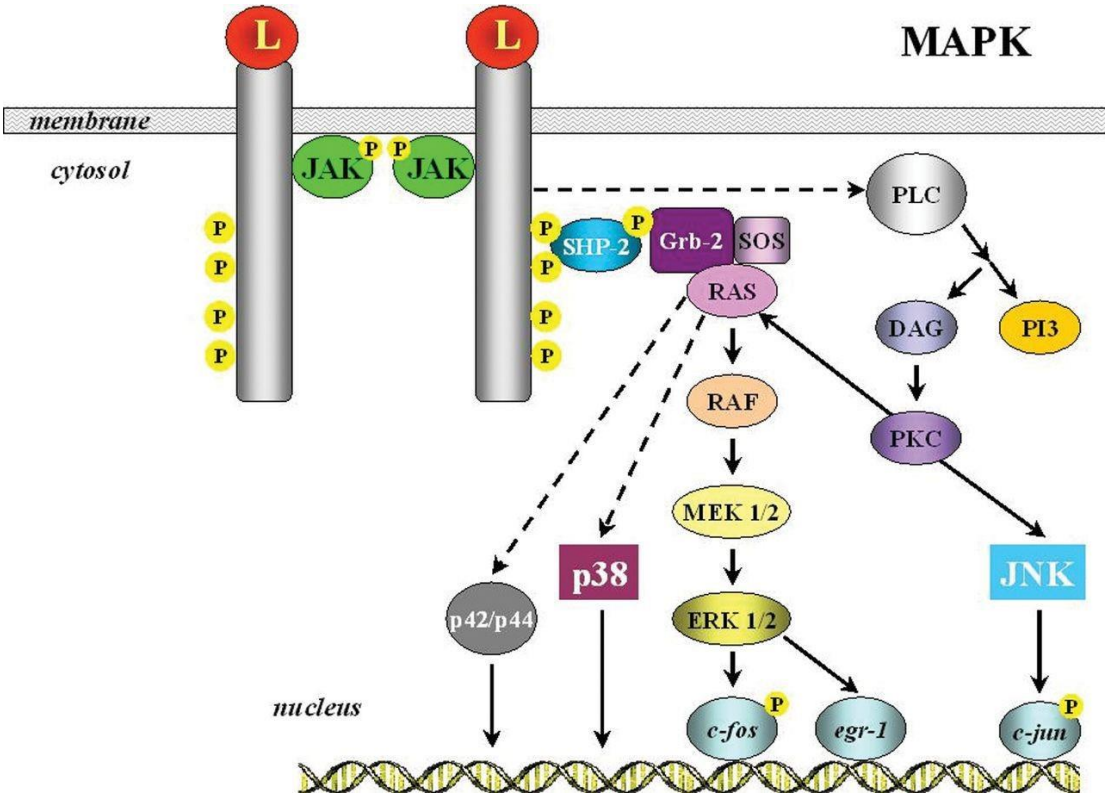



**Figure 1.6 SHP2-mediated signalling pathways.** SHP2 plays a central role in the transduction of mitogenic signals downstream from multiple growth factor (and cytokine) receptors. Ligand-mediated activation of receptor tyrosine kinase activity promotes sequential tyrosine phosphorylation of intracellular receptor tails. Phosphorylated tyrosine residues serve as docking sites for the engagement of SHP2 via its SH2 domains. This promotes activation of SHP2 by release of the autoinhibited conformation, allowing the propagation of intracellular signal transduction leading to transcriptional activation and expression of genes driving cellular proliferation, development and differentiation. Figure obtained from [102].

## 1.6 Aberrant SHP2 signalling in disease

### 1.6.1 Noonan syndrome, LEOPARD syndrome and leukaemias

In addition to the key regulation of signalling pathways controlling normal cellular physiology, dysregulation of SHP2 signalling has been frequently linked to a spectrum of myeloproliferative neoplasms and developmental disorders. In particular, germline mutations in the PTPN11 gene have been identified in approximately 40-50% of cases of NS. Despite clear links establishing the development of NS with defective PTPN11 gene function, gain of function mutations in *K-ras*, *Sos1* and *Raf1* genes have also been identified in NS patients devoid of PTPN11 mutations, strongly suggesting that dysregulation of the Ras/MAPK pathway is a key biochemical hallmark underlying this disorder [103]. Germline mutations in the PTPN11 gene have also been identified in LEOPARD syndrome (LS), a rarer but related condition to NS, with approximately 200 cases known to have been reported [104]. Somatic defects in the PTPN11 gene have been characterised in approximately 35% of JMML patient cases, and exhibit a reduced prevalence in related myeloproliferative disorders, such as myelodysplastic syndrome (MDS) [101]. Unfortunately, despite advances in our understanding of the genetic origins driving JMML, frontline therapies including stem cell transplantation provide only adequate symptomatic relief, with high relapse rates reported after treatment [105]. The identification of novel therapeutic targets driving JMML-like symptoms is therefore an urgent requirement for the effective control of this aggressive disease. Indeed, a deeper understanding of the biochemical consequences of JMML mutations in controlling SHP2 catalytic activity may offer a novel avenue for the development of more efficacious therapeutics with novel modes of action. At a biochemical level, the onset of NS is driven by the defective regulatory properties exhibited by mutant SHP2. More specifically, a

combination of structural, enzymatic and computational modelling studies suggest that these mutations, which are often found in the N-SH2 or PTP domain, exert their effect by destabilising the native auto-inhibitory mechanism of SHP2 leading to the aberrant control of functional phosphatase activity [106]. Similar mechanistic defects in SHP2 auto-inhibition are also exhibited by leukaemia-specific mutations located in the SHP2 N-SH2 domain leading to hyper-activation of MAPK signalling and marked cytokine sensitivity in myeloid-derived progenitor cells [107]. In contrast, LS-associated mutations typically reside in the phosphatase domain of SHP2 and manifest their effects by dramatically reducing phosphatase catalytic activity leading to an overall reduction in Ras/MAPK signalling [108]. This raises an interesting scientific paradox concerning how biochemically opposing mutations (e.g. activating mutations in NS versus inhibitory mutations in LS) ultimately lead to similar phenotypic observations. A number of potential explanations and key insights have been reported in an attempt to rationalise this apparent scientific quandary. One potential rationale suggests that overlapping Noonan and LEOPARD syndrome phenotypes may be driven by the propensity of mutant SHP2 to exert differential effects on Ras-MAPK signalling at temporally distinct phases of the developmental process [108]. In addition, structural insights from LS mutants reveal that despite exhibiting compromised catalytic activity, the presence of the mutation encourages longer lasting duration of a conformationally open state which is more amenable to interaction with upstream signalling activators and thus leads to an overall enhancement of catalytic activity [109, 110]. Alternatively, the capacity of SHP2 to perform PTP-independent functions as part of other signalling networks, or mutation-specific effects driving altered substrate recognition may also constitute a plausible biochemical rationale underlying the clinically similar phenotypes presented by Noonan and LEOPARD syndrome patients [72].

Although mutant SHP2 has been extensively linked to the onset of multiple myeloproliferative neoplasms, overexpression of wild-type SHP2 has also been observed as a pathological signature underlying chronic myelomonocytic leukaemia (CML) and systematic mastocytosis, both of which have been linked to oncogenic mutations in the *KIT* gene. Indeed, oncogenic activating mutations in the KIT receptor (specifically D816V mutations) stimulate aberrant activation of receptor signalling and promote hyperproliferation of bone marrow cells *in-vitro* and *in-vivo* leading to the onset of myeloproliferative disease [111, 112]. Interestingly, recent studies have identified an emerging role of SHP2 as a critical driver of oncogenic KIT activation in myeloproliferative disease [113]. These reports exposing the transforming ability of SHP2 in driving diverse myeloproliferative pathologies suggest that targeted SHP2 inhibition may provide a viable therapeutic avenue for the treatment of JMML and other myeloproliferative neoplasms driven by aberrant SHP2 expression or KIT hyper-activation [114].

### 1.6.2 Breast cancer

With an estimated 1.38 million cases documented annually worldwide, breast cancer represents the most frequently diagnosed cancer in women [115]. Like many cancers, breast cancer exhibits considerable inter-and intra-tumoural heterogeneity with a diverse mutational profile that acts to sustain the oncogenic phenotype [116]. Interestingly, although aberrant SHP2 signalling has been reported in breast cancer, observations of specific mutations in the PTPN11 gene have remained elusive. Indeed, recent studies have identified that up-regulation and overexpression of a key SHP2 docking protein acts a primary driver of SHP2-linked breast cancers [69]. In particular, the gene encoding the

SHP2 adaptor protein, GRB2-associated-binding protein 2 (Gab2), is amplified in approximately 10-15% of breast malignancies, with multiple studies linking up-regulation of SHP2 expression through amplified Gab2 as a key factor underlying the metastatic progression of breast cancer cells [117-119]. However, recent lines of evidence have demonstrated that SHP2 plays a central and previously unappreciated role in the initiation and maintenance breast tumour cells driving the metastatic phenotype underlying both human epidermal growth factor receptor 2 (HER2)-driven and triple negative breast tumours [120]. More specifically the study generated by Aceto *et al.* identified a distinctive genetic fingerprint that was consistently up-regulated in response to upstream activation of SHP2 signalling. Furthermore, this genetic signature was reported in approximately 50% of both primary breast malignancies and triple-negative breast cancers [117]. These key findings provide robust biological validation in redefining the therapeutic value of SHP2 as a potential oncology drug target and uncover novel mechanisms that could be exploited therapeutically as part of future breast cancer treatment strategies. In particular, patients harbouring tumours showing consistent up-regulation of this gene cluster in response to hyper-activated SHP2 could be stratified for treatment with a specific SHP2 inhibitor, thereby providing a possible therapeutic avenue for overcoming the resistant phenotype that frequently undermines current therapeutic regimes [121].

### 1.6.3 Gastric cancer

The infectious agent *Helicobacter pylori* has been suggested to underlie approximately 5% of all cancer cases in the UK [122], with around 30% of all stomach cancer cases linked to the primary etiological role of this pathogen in gastric carcinoma. More specifically, the onset of gastric carcinoma which represents one of the most commonly presented forms of

stomach cancer has been increasingly linked to cytotoxin-associated gene A (CagA)-specific strains of *Helicobacter pylori.* In particular, these strains exert their pathogenicity by injecting CagA into cells of the gastric epithelia via the evolutionarily conserved bacterial type IV secretion system, which functions as a "molecular syringe" by attaching to and puncturing the target cell membrane and injecting the CagA secretory factor directly into the cells. Following injection into cells of the gastric epithelia, CagA localises to the inner cytosolic leaflet of the plasma membrane where it undergoes phosphorylation on a number of key tyrosine residues by Src kinase family members Fyn and Lyn [123]. Tyrosine phosphorylation of CagA creates docking sites which invokes recruitment of SHP2 through binding of its tandem SH2 domains. This triggers constitutive activation of Ras-MAPK signalling leading to uncontrolled cellular proliferation and transformation. Therefore, the functional activation of SHP2 via its direct engagement of tyrosine phosphorylated CagA recapitulates the deregulated signalling responses exhibited by hyperactivated SHP2 signalling in multiple different cancers [120] (Figure 1.7).

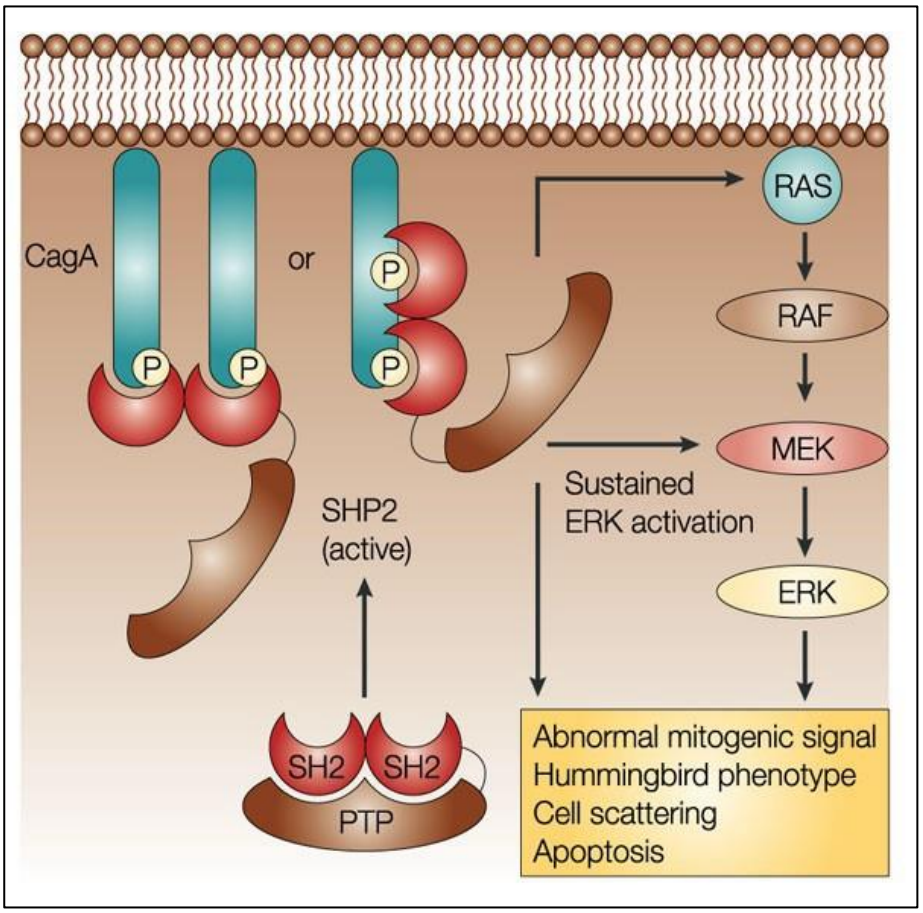



**Figure 1.7 CagA-mediated hyper-activation of SHP2 signalling leads to sustained Ras activation.** CagA is injected into gut epithelial cells via the Bacterial type IV secretion system of Helicobacter Pylori. CagA localises to the cytosolic leaflet of the plasma membrane and undergoes tyrosine phosphorylation by Src family kinases. The newly created phosphotyrosine residues act as docking sites for SHP2 recognition via its SH2 domains. This leads to hyperactivated SHP2 signalling and sustained Ras activation. CagA therefore serves to functionally mimic the gain of function mutation phenotype exhibited by hyper-activated SHP2. Figure taken from [124].

## 1.7 Protein tyrosine phosphatases as therapeutic targets

### 1.7.1 Modulation of phosphorylation signalling as a therapeutic strategy

Perturbing the balance of phosphorylation signalling results in compromised regulation of signal transduction networks, and serves as a biochemical hallmark underlying a range of critical human diseases. In this way, selective and functional modulation of phosphorylation-based signalling networks holds considerable therapeutic appeal. Indeed, many of the frontline drugs developed to specifically modulate protein tyrosine kinase function have entered the clinic and represent outstanding breakthroughs in the field of cancer therapeutics. In particular, the development of Gleevec (STI-571), a potent and targeted inhibitor of the p210 BCR-ABL oncogenic protein tyrosine kinase (PTK) was a significant therapeutic breakthrough for the treatment of CML [125]. Alongside small molecule inhibitor development, antibody-based therapies have also emerged as key therapeutic candidates, most notably Herceptin (Trastuzumab) which targets overexpression of the HER2 PTK receptor and is frequently overexpressed in approximately 25% of breast malignancies [126]. Despite advances in the range of therapeutic options available for the treatment of these human cancers, the development of chemotherapeutic resistance in patients administered with these treatments poses a robust barrier to the long-term treatment of these conditions [127, 128]. This suggests that identifying alternative approaches to targeting tyrosine phosphorylation based signalling pathways may have profound implications for the future development of novel therapeutics.

Over the past decade, considerable efforts have been devoted to the development of therapeutic candidates which selectively modulate PTK function. Although this has resulted in great success, and has led to a surge in the approval of PTK inhibitors for

clinical development [129], other equally important components of phosphorylation dependant signalling pathways, such as PTPs, have remained largely unexplored. The process of protein phosphorylation is a fundamentally reversible process that is critically dependant on the synergistic and competing activities of PTKs and PTPs, both of which function to maintain the cellular phosphoproteome. Given the multiple successful drug discovery initiatives targeting PTKs, and that the maintenance of homeostatic levels of protein tyrosine phosphorylation are governed by the highly co-ordinated activities of PTK and PTPs, PTPs could represent a class of signalling enzymes with considerable therapeutic potential.[13, 74].

### 1.7.2 The "undruggable" nature of protein tyrosine phosphatases

PTPs comprise a family of 107 signalling enzymes and play a central role in catalysing the dephosphorylation of phosphotyrosine-containing substrates [13]. Despite PTPs being considered as possible therapeutic targets, and the publication of several phosphatase inhibitors, phosphatases have been largely posited as an "undruggable" class of drug targets [130]. The proposed undruggability of phosphatases particularly refers to more conventional methods of drug discovery involving the development of traditional active-site directed inhibitors. For phosphatases, these conventional active-site directed approaches to drug discovery are often intractable. In particular, targeting phosphatases for inhibitor development presents several technical challenges which arise principally from the highly conserved and shallow active site architecture, the highly polar catalytic pocket, and the catalytically critical and redox-active cysteine. Owing to these challenges, the development of potent, selective and cell-active phosphatase inhibitors has been fraught with difficulties, often due to the charges and high reactivity of the final optimised ligands

that render them poorly cell permeable and of inadequate selectivity. This has led to suggestions that targeting phosphatases through active site-directed mechanisms represents an intractable approach and that alternative approaches centred on the identification and discovery of allosteric pockets may yield greater therapeutic value [131].

### 1.7.3 Current landscape of SHP2 inhibitors

Since its discovery as the first oncogenic PTP [74], the past decade has witnessed a concerted effort towards the development of potent and selective inhibitors for modulation of SHP2 function. Indeed, the development of high quality chemical probes is an exciting and rapidly growing area of drug discovery [132]. In the context of SHP2, the need for potent, selective and cell permeable inhibitors is urgently required both for functional annotation and exploration of SHP2-specific signalling pathways, and for enhancing the biological validation of SHP2 as an oncology therapeutic target.

In the infancy of SHP2 inhibitor development, one of the earliest reported compounds with demonstrable activity against SHP2 was the organometallic-based compound sodium stibogluconate (Compound 1; Figure 1.8; Table 1.2). Despite being used clinically for over 40 years as a treatment for the infectious parasitic-born disease, leishmaniasis, the precise mechanism of action remains unclear [133]. Interestingly, it is now known that stibogluconate exerts its inhibitory activity through the potent and irreversible inhibition of SHP1, SHP2 and PTP1B *in vitro*, which elicits activation of a broad spectrum of cytokine signalling cascades [134]. Further interrogation of the mechanistic basis of stibogluconate action revealed potent anti-tumour activity mediated through IFN-α2b activation in an *in vivo* mouse model. This has now led to initiation of clinical trials investigating the efficacy of stibogluconate for the treatment of melanoma and several other human cancers [135].

However, administration of stibogluconate is now widely known to elicit multiple harmful side effects including cardio and hepatotoxicity. Despite these undesirable side effects, the design of alternative chemical inhibitors with novel scaffolds could still hold therapeutic benefit for the treatment of various cancers [133].

Screening of the chemical library belonging to the National Cancer Institute (NCI) diversity set led to the observation of inhibitor NSC-87877 (Compound 2; Figure 1.8; Table 1.2) [136]. NSC-87877 demonstrates reasonable potency *in vitro* with an $IC_{50}$ of 0.3 μM against SHP2, but lacks selectivity over the closest homolog of SHP2, SHP1. Despite the lack of SHP1 selectivity, NSC-87877 displays a reasonable selectivity profile over other PTPs and has also been reported as an effective inhibitor of DUSP26 signalling [102]. Interestingly, despite possessing a core disulphonic-acid based moiety which carries a high density of negative charge, NSC-87877 demonstrated in cell activity by inhibiting EGF-stimulated activation of SHP2 signalling in HEK293 cells at a concentration range of 10-50 μM. Although no clear structural evidence has been reported to unequivocally confirm the binding mode of NSC-87877, a combination of *in silico* modelling and mutational analysis suggested an active-site directed binding mode for this compound. Furthermore, *in vivo* analysis of NSC-87877 activity has demonstrated that inhibitor treatment is sufficient to reverse the development of experimental autoimmune encephalomyelitis in a mouse model of chronic relapsing-remitting multiple sclerosis [137]. Although NSC-87877 has reported demonstrable activity *in vivo,* whether the phenotypic effects observed are a result of specific inhibition of SHP2 or mediated through off target effects (i.e. through polypharmacology) remains unclear.

Interestingly screening the same NCI-diversity set chemical library also resulted in the identification of inhibitor NSC-117119 (Compound 3; Figure 1.8; Table 1.2). Compared to NSC-87877, NSC-117119 exhibits modest potency with an $IC_{50}$ = 47 μM, and also cross-inhibition of SHP1 and PTP1B. Similarly, although the existence of structural evidence confirming the binding mode of NSC-117119 has yet to be reported, modelling and ligand docking studies in conjunction with synthetic chemistry led to the evolution of two related chemotypes which both exhibited improved potency and selectivity. These include compound 14a (Compound 4; Figure 1.8; Table 1.2) which has a SHP2 $IC_{50}$ of 0.8 μM and exhibits approximately 15-fold greater selectivity over SHP1, and compound SPI-112 (Compound 5; Figure 1.8; Table 1.2) which has an $IC_{50}$ of 1 μM and exhibits approximately 18-fold selectivity over SHP1 and 14-fold selectivity over PTP1B [138]. Further modification of the previously cell-impermeable compound SPI-112 to the corresponding methyl ester analog SPI-112Me (Compound 6; Figure 1.8; Table 1.2) generated a SHP2 inhibitor with in-cell activity at concentrations of 20 μM. Screening of a natural product library also identified inhibitor 2a (Compound 7; Figure 1.8; Table 1.2) as a relatively potent inhibitor with an $IC_{50}$ = 2.5 μM and which also exhibited approximately 20-fold selectivity for SHP2 and PTP1B over a broad panel of other PTPs in a biochemical assay [139]. Interestingly, in the absence of a negatively charged anionic handle which is usually a common structural feature of many phosphatase inhibitors, it would be interesting to observe the performance of this inhibitor and establish its activity and cell permeability using a cell-based assay system.

Virtual screening of a 2.7 million compound collection against the SHP2 catalytic domain led to the identification of the inhibitor PHPS1 (Compound 8; Figure 1.8; Table 1.2) which exhibits an $IC_{50}$ of 0.7 μM with 15-fold selectivity over SHP1 and a panel of other

phosphatases. Although structural evidence confirming the binding mode of PHPS1 to the SHP2 catalytic domain remains elusive, *in silico* modelling suggested an active-site directed binding mode. In addition, PHPS1 exhibited in-cell activity through blocking HGF-induced ERK phosphorylation in a range of tumour cell lines [140] and has been shown to inhibit cardiac hypertrophy driven by hyperactivated gain of function SHP2 mutants in cultured cardiomyocytes [141].

One of the more well validated SHP2 inhibitors was reported by Zhang *et al.*, where screening of a combinatorial library using a salicylic acid based scaffold enabled the identification of inhibitor IIB08 (Compound 9; Figure 1.8; Table 1.2), which has an $IC_{50}$ of 5.5 μM against the SHP2 catalytic domain and a modest 3-fold selectivity over SHP1 and PTP1B. Kinetic analysis of compound IIB08 inhibition demonstrated that the compound is a reversible non-competitive inhibitor of the SHP2 catalytic domain. Furthermore, compound IIB08 demonstrates activity in a cell-based assay by blocking EGF-stimulated activation of ERK signalling leading to proliferation of haematopoietic cells in response to GM-CSF cytokine signalling. A thorough *in vivo* analysis of IIB08 activity demonstrated reversal of tumour xenograft growth in mouse models of non-small cell lung cancer [142]. Interestingly, compound IIB08 has also demonstrated activity in mast cell models of leukaemia where administration of the compound effectively reduced the growth of cells harbouring oncogenic forms of the protein tyrosine kinase KIT. The positive anti-myeloproliferative promoting effects of IIB08 led to extended survival of mice in models of leukaemia. In addition to the wealth of *in vitro* and cell-based activity demonstrated for compound IIB08, structural studies interrogating the mode of compound binding to the SHP2 catalytic domain have provided key insights into the interactions formed between active site residues of the enzyme and the salicylic acid moiety. This crystal structure

provides a basis for the rational design of more potent and efficacious inhibitors of the SHP2 catalytic domain based on the salicylic acid scaffold [143].

Multiple other SHP2 inhibitors have been identified through combinations of *in silico* docking, compound library screening and medicinal chemistry. For example, the oestrogenic-based compound estramustine phosphate (Compound 10; Figure 1.8; Table 1.2) which harbours a nitrogen mustard moiety and is used as a therapeutic agent in the treatment of prostate cancer is a SHP2 inhibitor with a relatively moderate potency of 17 μM [144]. Kinetic characterisation of estramustine phosphate suggests the compound displays mixed enzyme inhibition, with *in silico* docking analysis suggesting that the compound binds with an active-site directed mode of inhibition. Interestingly, the chemical scaffold of estramustine phosphate is structurally similar to the PTP1B inhibitor MSI-1436, which has recently been identified as an allosteric inhibitor of PTP1B that simultaneously targets the disordered C-terminus as well as an additional secondary pocket in close proximity to the active site [145]. Thus, although the structural resemblance to MSI-1436 may indicate a similar binding mechanism for inhibition of SHP2, this has yet to be established.

More recently, natural product inhibitors have been identified as specific inhibitors of the SHP2 catalytic domain. In particular, the compound cryptotanshinone (Compound 11; Figure 1.8; Table 1.2) derived from the *Salvia miltiorrhiza* plant [146] was observed to inhibit both SHP1 and SHP2 with a similar $IC_{50}$ value of 23 μM. Cryptotanshinone also exhibits cell-based activity through effective repression of SHP2 activated IL-3 signalling. Profiling the activity of cryptotanshinone in an enzymatic assay revealed an irreversible mechanism of inhibition, which has been linked to generic mechanisms identified for

quinone-based phosphatase inhibitors that include oxidation of the redox sensitive catalytic cysteine residue of PTPs [147]. Despite exhibiting promising activity in cell-based assays, the reactive oxygen species generated as a by-product of inhibition would be likely to induce toxicity and suggest that side effects of this nature would be difficult to avoid if these classes of inhibitors were pursued for further development of more potent SHP2 inhibitors. Another natural product inhibitor demonstrating specific inhibition of the SHP2 catalytic domain is tautomycetin (Compound 12; Figure 1.8; Table 1.2), which exhibits potent inhibition of serine-threonine phosphatases PP1 and PP2A, with $IC_{50}$ values of 1.6 nM and 62 nM, respectively. Recently, tautomycetin was discovered as an *in vitro* active SHP2 inhibitor with an $IC_{50}$ of 2.9 μM and 5-fold selectivity over SHP1 [148]. A crystal structure of tautomycetin bound to the SHP2 catalytic domain revealed that the inhibitor binding mode is analogous to phosphotyrosine containing substrates with the lipophilic anionic head group forming the dominant interaction with the active site [145].

Despite considerable efforts directed towards the development of potent and selective PTP inhibitors, many of the currently available compounds lack the potency, selectivity and cell permeability required for use in animal studies. However, recently a group from Novartis reported the identification of a highly specific, potent and orally bioavailable SHP2 inhibitor that operates through an allosteric mechanism of inhibition [149]. The allosteric SHP2 inhibitor, known as SHP099 (Compound 13; Figure 1.8. Table 1.2), was developed through an intricate series of screening assays which focused on the identification of small molecules which repressed activation of full length SHP2 by a pTyr-based IRS peptide, but had no inhibitory effect on the isolated catalytic domain. The crystal structure of SHP099 bound to full length SHP2 revealed a previously undisclosed tunnel-like binding site located between the catalytic, N-and C-SH2 domains. This suggested that SHP099

exerts its inhibitory effect by locking full length SHP2 in the auto-inhibited conformation and therefore exploits the natural inhibitory mechanism of the enzyme. Furthermore, SHP099 is sufficiently potent ($K_D$ = 71 nM), and exhibits excellent selectivity with no significant activity against a broad panel of other PTPs and kinases. However, despite the advantageous properties of this inhibitor, concerns have been raised suggesting that the authors failed to explore the potential immune-stimulatory effects that might ensue as a result of blocking the immune-inhibitory response normally fine-tuned by uninhibited SHP2 [150].

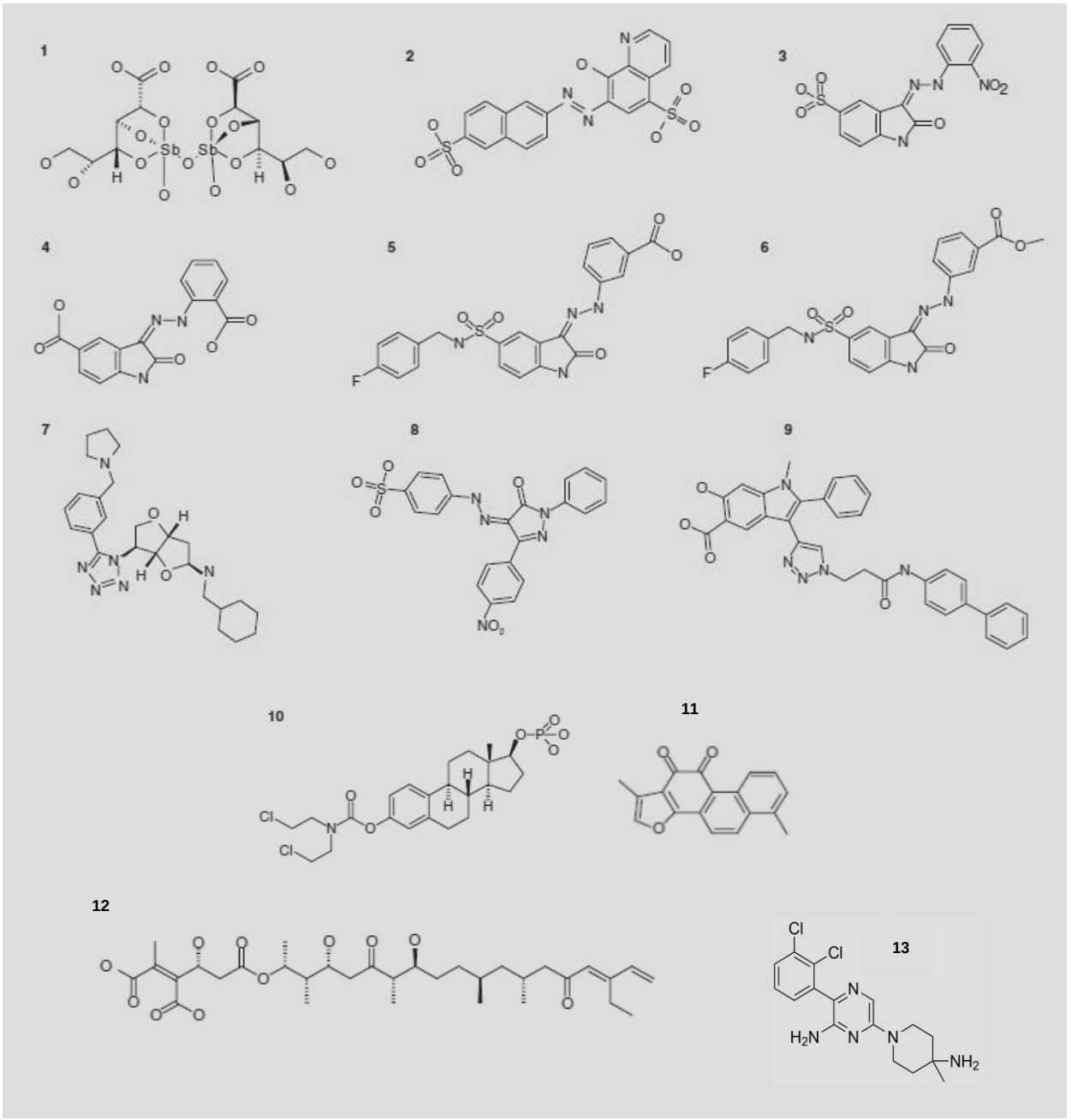



**Figure 1.8 Chemical structures of SHP2 inhibitors. (**1) Sodium Stibogluconate. (2) NSC-87877. (3) NSC-117119. (4) Compound 14a. (5) SPI-112. (6) SPI-112Me. (7) Compound 2a. (8) PHPS1. (9) IIB08. (10) Estramustine phosphate. (11) Cryptotanshinone. (12) Tautomycetin. (13). SHP099.

| Compound number | Name | $IC_{50}$ or $K_i$ (SHP2) | Selectivity | Activity (cell based; *in-vivo*) | Mechanism (type of inhibition) |
|---|---|---|---|---|---|
| **1** | Sodium stibogluconate | 100 µg/ml | SHP1: 10 µg/ml<br>PTP1B: 10 µg/ml | Yes; yes | Covalent |
| **2** | NSC-87877 | 0.3 µM (PTP domain)<br>1.3 µM (full length SHP2) | SHP1: 1-fold<br>PTP1B: 5-fold<br>HePTP: 24-fold<br>Others: >200-fold<br>DUSP26: 17 µM | Yes; yes | Competitive |
| **3** | NSC-117119 | 47 µM | SHP1: 68 µM<br>PTP1B: 96.7 µM | Inactive; NA | NA |
| **4** | Compound 14a | 0.8 µM | SHP1: 15.4 µM<br>PTP1B: 1.5 µM | Inactive; NA | NA |
| **5** | SPI-112 | 1 µM | SHP1: 18.3 µM<br>PTP1B: 14.5 µM | Inactive; NA | NA |
| **6** | SPI-112Me | >100 µM (*in-vitro*)<br>~20 µM (cell based assays) | NA (see SPI-112 data) | Yes; NA | NA |
| **7** | Compound 2a | 2.5 µM | PTP1B: >40-fold | NA | NA |
| **8** | PHPS1 | 0.7 µM | SHP1: 8-fold<br>PTP1B: 15-fold<br>Others: 25-fold | Yes; NA | Competitive |
| **9** | IIB08 | 5.5 µM | SHP1: 15.7 µM<br>PTP1B: 14.3 µM<br>LYP: 25 µM<br>FAP1: 20 µM<br>CD45: 30 µM<br>LMWPTP: 31 µM | Yes; yes | Reversible and non-competitive |
| **10** | Estramustine phosphate | 17 µM | SHP1: 40 µM<br>PTP1B: 62 µM<br>HePTP: 153 µM | NA; yes | Mixed |
| **11** | Cryptotanshinone | 22.5 µM | SHP1: 40 µM<br>PTP1B: 34 µM<br>CD45: 42 µM<br>LAR: 37 µM<br>MEG2: 59 µM<br>TCPTP: 56 µM | Yes; NA | Oxidation of catalytic Cys and alkylation |

| | | | | | |
|---|---|---|---|---|---|
| **12** | Tautomycetin | 2.9 μM | SHP1: 14.6 μM<br>PTP1B: 41 μM<br>LYP: 20 μM<br>Others: >50 μM | Yes; yes | NA |
| **13** | SHP099 | 0.071 μM | SHP1: >100-fold | Yes; yes | Allosteric |

**Table 1.2 Pharmacological profiles of selected SHP2 inhibitors.** Pharmacological characterisation of SHP2 inhibitors. The $IC_{50}$ value is defined as the inhibitor concentration eliciting a 50% reduction in enzymatic activity. The inhibitor constant, $K_i$, provides an indication of inhibitor potency and is also defined as the inhibitor concentration required to produce half-maximum inhibition. Although $IC_{50}$ and $K_i$ values both represent measurements of inhibitor potency, the conditions under which these measurements are obtained is fundamentally different. Whereas $K_i$ values are measured *in-vitro* against isolated and purified enzyme, inhibitor $IC_{50}$ values are recorded *in-vivo* where enzyme inhibition is measured in the presence of additional factors such as competing macromolecules and molecular crowding. As such, $IC_{50}$ and $K_i$ values are not directly comparable, with $IC_{50}$ values yielding more representative derivations of inhibitor activity under physiologically relevant conditions. Inhibitor selectivity refers to the property determined by combining the affinity of the molecule at a particular binding site. The activity of an inhibitor refers to the ability to inhibit a biochemical reaction by a known quantity (in a biochemical assay), or reverse a phenotypic characteristic if activity is measured in an *in vivo* context. The mechanism of an inhibitor refers to the type of enzymatic inhibition displayed by a particular molecule and specifically how the inhibitor interferes with the kinetics of enzyme mechanisms and substrate turnover. Such mechanisms include competitive, non-competitive and uncompetitive modes of inhibition.

## 1.8 Screening approaches for phosphatase inhibitor discovery

### 1.8.1 High-throughput compound screening for phosphatase inhibitor discovery

Although a successful approach for the development and progression of clinical agents when applied to alternative enzyme systems (such as the kinases family of enzymes), high-throughput screening efforts to develop cell-active phosphatase inhibitors with attractive selectivity has, until recently [149, 151], been met with comparatively modest success. In particular, this disproportionate success has been underscored by the high incidence of false positive hits emerging from phosphatase screening campaigns. Although a precise explanation for this observation has yet to emerge, multiple mechanisms underlying such phenomena have been suggested. These include: 1) the propensity for phosphatase inhibitors to form micelle-like aggregates that non-specifically inhibit enzyme activity [152], 2) the high reactivity of the nucleophilic active site cysteine [153], and 3) the redox sensitivity of the active site cysteine leading to irreversible inactivation of phosphatase activity [154, 155]. In addition to the technical challenges posed by the phosphatase active site chemistry and reactivity, phosphatases are renowned for a notable lack of obviously druggable surface pockets that could otherwise be exploited for inhibitor development. Instead, phosphatases harbour very shallow surface pockets which typically render these enzymes intractable to conventional high-throughput screening. This is primarily due to the enhanced molecular size of the compounds used in conventional high-throughput screens which hinders the accessibility of these compounds for binding these pockets.

Despite the absence of clinically efficacious phosphatase inhibitors, a combination of conventional high-throughput screening coupled with fragment-based linking approaches have been successful in generating novel, potent and highly selective inhibitors of the

protein tyrosine phosphatase, PTP1B [156, 157]. The development of these inhibitors greatly benefited from the versatility of NMR spectroscopy which proved to be particularly useful for enabling the detection of a diverse range of ligand affinities and thus offered greater sensitivity for the screening and validation of early stage compound hits against PTP1B. The generation of these inhibitors required HSQC NMR screening of uniformly $^{15}N$-labelled PTP1B against a 10,000 compound library with hits validated in the first instance by assessment of chemical shift perturbations in $^{15}N$-$^{1}H$-resolved HSQC spectra. Previously determined NMR backbone assignments of the PTP1B protein then enabled chemical shift perturbations to be mapped to key residues V49, G228 and G218, all of which were located in and around the PTP1B catalytic site. Analysis of the initial screening data revealed a phosphotyrosine mimetic with a $K_D$ of 100 μM which was subsequently optimized into a more potent active site ligand with a $K_D$ of 26 μM. The x-ray crystal structure of the optimised phosphotyrosine mimetic in complex with PTP1B provided a robust model of ligand binding and confirmed the molecular basis of active site specificity of the bound ligand. Unfortunately, the highly conserved architecture and chemistry of the phosphatase active site poses significant obstacles for the development of conventional active-site directed inhibitors. This in turn also poses a significant challenge for achieving the selectivity required for an efficacious clinical inhibitor. Circumventing these difficulties requires the development of a re-directed strategy involving the synthesis of bivalent inhibitors capable of forming interactions in close proximity to and outside of the active site cavity. The development of bivalent inhibitors offers the advantage of multiple contact points with a particular target of interest each possessing differential potency and specificity which ultimately enhances inhibitor selectivity and affinity.

In the same study Szczepankiewicz et al., were successful in discovering inhibitors binding to a secondary binding site which was achieved through the utilization of a second screening approach involving the selective labelling of PTP1B with $^{13}$C-methionine in order to observe chemical shift perturbations of 13C-methionine resonances in $^{13}$C-$^{1}$H HSQC spectra. This screening approach was successful in identifying a ligand bound to this site which through optimization and iterative cycles of synthetic chemistry was conjugated to the initial active site ligand using a chemical linker. The final inhibitor exhibited impressive activity in a pNPP biochemical assay with a $K_i$ of 20 nM, along with an equally impressive specificity profile against a panel of phosphatases including leukocyte antigen-related tyrosine phosphatase (LAR), SHP2, (protein tyrosine phosphatase, receptor type, C) PTPRC and calcineurin ranging from 36-fold to 10,000 fold selectively with a moderate two-old selectivity observed over T-cell tyrosine phosphatase (TCPTP). This approach established a clear paradigm for phosphatase inhibitor discovery, suggesting that the development of bivalent inhibitors provides considerable benefits which bolster the affinity and specificity of the final molecule.

Despite the transformative nature of this strategy in the context of phosphatase drug discovery, the final PTP1B bivalent inhibitors developed through this approach exhibited poor physicochemical properties and unfavourable bioavailability, despite demonstrating promising selectivity [156]. However, alternative approaches utilising NMR fragment-based screening methods and x-ray crystallography have been successful in generating PTP1B inhibitors with demonstrable cell activity [157]. These screening approaches centred heavily on the use of monocarboxylic and non-carboxylic acid based fragments as the precursor chemotypes which were then optimised into more potent catalytic site ligands. Such chemotypes were purposefully selected on the basis of their low charge density in order to

maximise the possibility of generating a cell permeable inhibitor. The structure-guided modification of the catalytic site ligands and a second phosphotyrosine binding pocket yielded a potent and selective PTP1B inhibitor with improved cell permeability that vastly superseded the bivalent PTP1B inhibitors developed by the same group [156]. The final inhibitors yielded low micromolar affinity and greater than 30-fold selectivity over related phosphatases, including TCPTP [158].

More recently, the application of high-throughput screening has also been successfully applied to the discovery of allosteric phosphatase inhibitors. For the past decade the design of allosteric phosphatase inhibitors has held considerable appeal, primarily due to the frustrating technical challenges of developing selective phosphatase inhibitors targeting the active site. This has led to the successful development of selective allosteric inhibitors targeting oncogenic phosphatases Wip1 [159] and Eya2 [160]. More recently, a team from Novartis reported the development of a potent, selective and orally bioavailable inhibitor, SHP099, for the protein tyrosine phosphatase SHP2, which appears to operate through an allosteric mechanism of inhibition. More specifically, the inhibitor exploits the unique auto-regulatory mechanism of SHP2 by binding to a centrally located tunnel away from the phosphatase active site. In the inhibitor-bound conformation, SHP2 is locked in the auto-inhibited state and is thus refractory to activation. The development of SHP099 was made possible through an intricately designed high-throughput screening strategy directed solely towards the development of allosteric inhibitors. A key factor influencing the design of this strategy centres on the observation that SHP2 is activated through the binding of doubly phosphorylated peptide and proteins through engagement of N-SH2 and C-SH2 domains which triggers release of auto-inhibition and the opening of the active site to facilitate substrate turnover [80, 161]. To discover inhibitors capable of exploiting this auto-

inhibitory mechanism, the Novartis researchers screened a library of 100,000 compounds at a single concentration of 20 μM against full length SHP2 that had been partially activated by equilibration with 0.5 μM of a bisphosphorylated IRS-1 peptide. From the primary screen, a total of 900 compounds were found to inhibit full length SHP2 by 30% or greater. These compounds were then progressed to assess inhibition in three distinct biochemical assays: (1) using the catalytic domain of SHP2 only, (2) assaying SHP2 activity in the presence of partially activating concentrations of bisphosphorylated IRS-1 peptide, and (3) assaying SHP2 activity in the presence of fully activating levels of bisphosphorylated IRS-1 peptide. Compounds targeting the SHP2 catalytic domain only were prematurely excluded from further screening to identify allosteric inhibitors. Profiling of the inhibitors tested in each of these assays identified six inhibitors demonstrating no inhibition against the SHP2 catalytic domain, moderate inhibition against full length SHP2 equilibrated with 0.5 μM of a bisphosphorylated IRS-1 peptide, and reduced inhibition against full length SHP2 equilibrated with 5 μM of a bisphosphorylated IRS-1 peptide. The inhibitor displaying the most prominent inhibition of full length SHP2 (referred to as SHP836) was chemically optimised into the inhibitor SHP099 which exhibited >70-fold improvement in potency yielding a final $IC_{50}$ = 71 nM. SHP099 also demonstrated a remarkable selectivity profile showing no detectable activity against a panel of 21 phosphatases and 66 kinases [162].

### 1.8.2 Fragment-based screening for phosphatase inhibitor discovery

In addition to the use of high-throughput screening for the development of phosphatase inhibitors, fragment-based screening approaches have also been employed to explore novel chemical scaffolds as starting points for phosphatase inhibitor development. More

recently, fragment-based screening efforts have been utilised to develop inhibitors against the dual-specificity phosphatase CDC25B [163]. To discover novel CDC25B inhibitors a diverse library of 1500 chemical fragments were screened by $^{15}N$-$^{1}H$ NMR HSQC at a single concentration of 250 μM. Profiling of the fragment hits identified a single hit yielding chemical shift perturbations which when mapped to the previously determined backbone assignments disclosed a binding site situated 15Å away from the active site. An x-ray crystal structure revealed the fragment binding location occupied a small pocket in close proximity to the CDK2 substrate binding site. The researchers also noticed a bound sulphate ion (from the crystallization buffer) in the same crystal structure which they attempted to link to the fragment hit. However, although the optimised molecule demonstrated inhibitory activity in a biochemical assay, the final potency was moderate with an $IC_{50}$ = 1-2 mM.

The application of fragment-based screening methods to develop selective phosphatase inhibitors remains a significantly underexplored area of research. Despite the obvious advantages of fragment-based screening for identifying novel druggable pockets, the results of the CDC25B study suggest that large and chemically diverse libraries containing >1500 fragments are likely to be the minimum requirements for the screening of phosphatases. The low hit rates observed in fragment-based screens typically correlate with the druggable potential of a particular target, and in this instance underscore the significant challenge of developing selective phosphatase inhibitors using such an approach [164]. Although high-throughput screening has successfully generated cell-active phosphatase inhibitors, these screens have greatly benefited from intricately designed screening cascades and innovative chemical linking strategies which were precisely tailored to exploit the selection of inhibitors with discrete modes of inhibition.

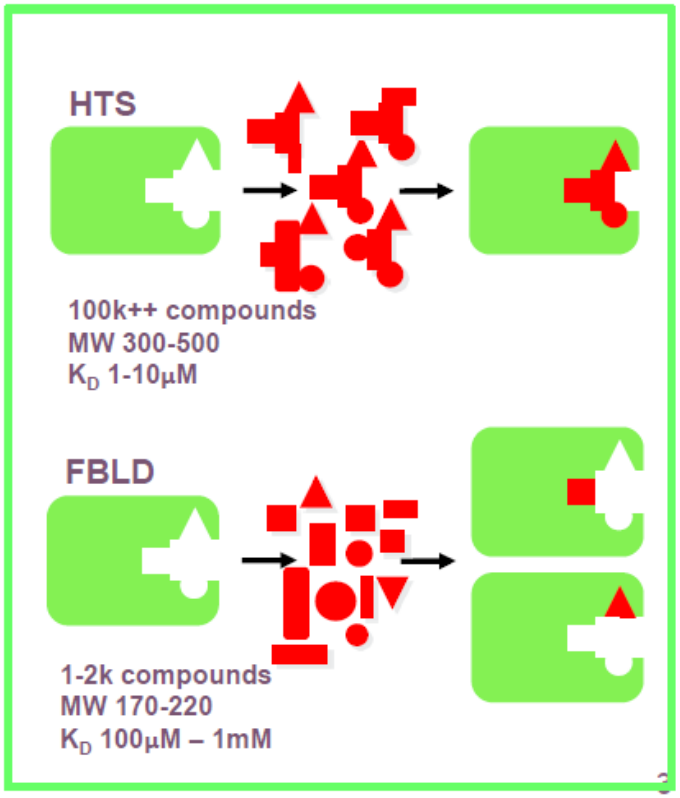



**Figure 1.9 High throughput screening vs Fragment-based ligand discovery.** High-throughput screening (HTS) typically involves screening large libraries of chemically advanced scaffolds (usually 100,000 – 1,000,000 compounds) with molecular weights in the range of 300 Da – 500 Da. However, fragment-based ligand discovery involves screening a comparatively smaller library (usually between 1,000 – 2,000 fragments) of low molecular weight fragments in the range of 150 Da - 250 Da. Employing fragment-based ligand discovery offers the advantage of screening smaller libraries due to the ability of fragments to sample large amounts of chemical space.

### 1.8.3 Fragment-based ligand discovery

Fragment-based ligand discovery (FBLD) has emerged as a leading approach for accelerating the early progression of small molecule therapeutics into approved clinical agents [165-167]. Over the past decade, the technique of FBLD has been routinely applied in both academic and industrial arenas and has provided clear strategies for stimulating the development of small molecule ligands, guiding structure-based drug discovery efforts, and revaluating approaches to high-throughput screening [168-170]. The fundamental rationale underpinning the FBLD approach is based on a number of key principles. The first of these principles is linked to the idea that small numbers of low molecular weight compounds (termed "fragments") can sample a large population of chemical space [171]. The second of these principles refers to the fact that as the molecular weight of a molecule increases along with an increase in molecular complexity, the likelihood of an unfavourable interaction occurring will be greatly increased [172]. Thus, by virtue of their low molecular complexity, small molecular weight compounds such as fragments are inherently less selective than compounds with larger molecular weights and are therefore more capable of interacting with a larger number of protein targets than more conventional drug-sized compounds [173]. The ability of small molecular fragments to sample chemical space efficiently also has profound implications for the size of chemical libraries for high-throughput screening cascades. Typically, conventional high-throughput screening cascades employ chemical libraries containing between 100,000 – 1, 000, 000 compounds, whereas libraries utilised for fragment-based ligand discovery contain a significantly lower number of molecular fragments usually in the range of 500-2000 compounds. Other than number of compounds, the key difference between chemical libraries used for conventional high-throughput screening and fragment screening libraries concerns the

molecular weight of the constituent chemotypes, along with their predicted affinities. The compounds contained in conventional high-throughput screening libraries typically occupy molecular weights between 350 Da – 500 Da, with the expected affinity ranges for these compounds emerging from high-throughput assays ranging from the nanomolar to low micromolar potency. In contrast however, compounds selected for fragment library screening are typically low molecular weight compounds with molecular weights in the range of 100 Da - 300 Da with affinities ranging from micrormolar to millimolar potency [174]. The third guiding principle of the FBLD approach is that despite the relatively low affinity of small molecule fragments, these chemotypes can be chemically optimised and expanded through iterative medicinal chemistry into more potent high affinity ligands. This iterative cycle of fragment evolution is greatly aided by structural insights obtained from x-ray crystal structures of fragments bounds to their protein targets.

The low molecular weight of chemical fragments often results in low affinity binding interactions due to the smaller number of molecular contacts formed between the fragment scaffold and the protein target. Detection of such low affinity interactions therefore requires highly sensitive and versatile screening assays in order to provide robust confirmation and validation of genuine fragment hits. More commonly, fragment library screening assays typically employ biophysical methods as the primary screening methodologies, and are often conducted in parallel using two or more independent methods in order to cross validate hits and reduce false positive hit rates. The range of suitable biophysical methodologies for fragment library screening is vast and includes NMR methods [175-178], X-ray crystallographic screening [179, 180], isothermal titration calorimetry [181], protein thermal shift [182], affinity capillary electrophoresis [183], weak affinity chromatography [184] and SPR [185]

## 1.9 Theoretical basis of experimental techniques

### 1.9.1 Nuclear Magnetic Resonance Spectroscopy

#### 1.9.1.1 The physical basis of NMR spectroscopy

Nuclear magnetic resonance spectroscopy is a powerful experimental tool facilitating the analysis of molecular structure in the solution state allowing interrogation of the interactions between individual atomic nuclei to probe their spatial relationships and dynamic propensities. In order to achieve this, NMR relies principally on exploiting the differential magnetic properties of certain atomic nuclei in the presence of an external magnetic field.

In their simplest form, atomic nuclei can be imagined as miniature bar magnets with north and south poles aligned at opposite ends along a vertical axis. An intrinsic characteristic of many atomic nuclei concerns a property known as nuclear spin. This nuclear spin can be thought of as a localised magnetic field precessing about a central plane in a direction perpendicular to the vertical north-south axis. The spin of an individual nucleus precessing in this manner is thus referred to as a magnetic moment (or angular momentum) which, by convention, contains two fundamental properties: a direction and a magnitude. By virtue of containing these two components, nuclear spin is thus operationally defined as a vector quantity. In many atomic nuclei (such as $^{12}C$), the spins of subatomic particles (protons, neutrons and electrons) are paired in an anti-parallel fashion, such that the nucleus has no overall spin. However, for some atoms (such as $^{1}H$, $^{13}C$ and $^{15}N$) the nucleus does possess an overall spin and is therefore a property that can be exploited in NMR experiments. The principles for determining the overall spin of an atomic nucleus are as follows:

1. If the number of neutrons and the number of protons are both even, the nucleus will have no overall spin
2. If the sum of the number of neutrons and the number of protons is an odd number, then the nucleus will have a half-integer spin (e.g. ½)
3. If the number of neutrons and the number of protons are both odd then the nucleus will have an integer spin (e.g. 1, 2, 3)

According to the elementary principles of quantum mechanics, a nucleus of spin $I$ can be quantified as having $2I + 1$ possible orientations. Therefore, in the case of nuclei with spin ½ (such as $^{1}H$, $^{13}C$ and $^{15}N$), quantum mechanics states that these nuclei can exist in two different energy states: + ½ and - ½. In the absence of an external magnetic field, these two orientations of the nuclear angular momentum will be equal in energy. However, in the presence of an external magnetic field the + ½ and - ½ nuclei will be differentially split into two different spin populations based on the orientation of the individual nuclear spins with respect to the external magnetic field. Those nuclei with spin + ½ aligning with the direction of the external magnetic field will be stabilised and thus lowered in energy. Conversely, those nuclei with spin - ½ which oppose the direction of the external magnetic field will be destabilised and thus excited to higher energy levels. In this way, the differential magnetic properties of + ½ and - ½ nuclei in the presence of an externally applied magnetic field creates an energy difference between the two nuclei populations which is proportional to the strength of the external magnetic field (Figure 1.10).

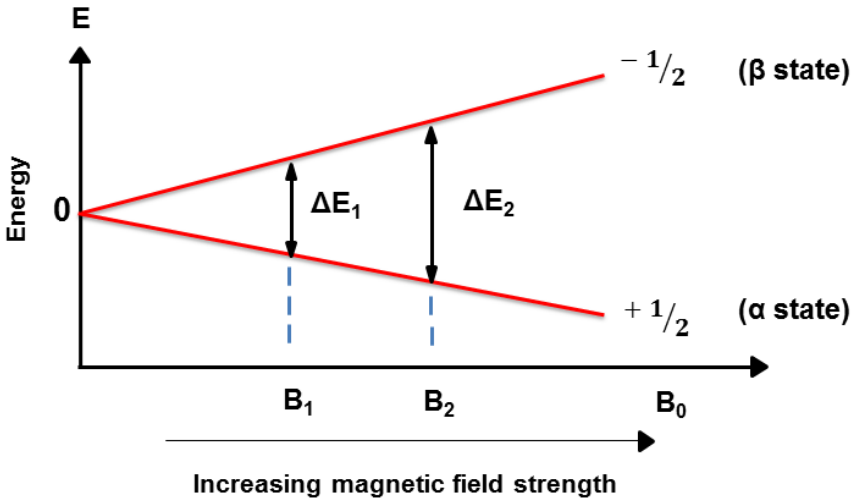



**Figure 1.10 Splitting of spin ½ nuclei in an external magnetic field.** According to quantum mechanics, a nucleus of spin *I* can be assigned 2*I* + 1 possible orientations. This asserts that nuclei with spin ½ (such as $^{1}$H, $^{13}$C and $^{15}$N) can occupy two different energy states: + ½ and - ½. In the absence of an external magnetic field, these two orientations of the nuclear angular momentum will be equal in energy. However, in the presence of an external magnetic field the + ½ and - ½ nuclei will be split to form two different spin populations. Nuclei with spin + ½ aligning with the direction of the external magnetic field (α state) will be stabilised and lowered in energy. Conversely, nuclei with spin - ½ which oppose the direction of the external magnetic field (β state) will be destabilised and promoted to higher energy levels. The differential magnetic properties of + ½ and - ½ nuclei in the presence of an externally applied magnetic field creates an energy difference between the two nuclei populations which is proportional to the strength of the external magnetic field.

Briefly, in order to generate an NMR signal the samples containing + ½ and - ½ nuclei populations in the presence of an external magnetic field are irradiated by applying a radiofrequency pulse with a frequency proportional to the energy difference between the two nuclei populations. The application of a radiofrequency pulse triggers the excitation of + ½ nuclei from the lower energy state to the higher energy state occupying - ½ nuclei. During application of the radiofrequency pulse, the net magnetisation vector precessing around the Z dimension is tipped into the XY plane. The precession of the net magnetisation vector in the XY plane then generates an oscillatory signal which induces a current in the coil surrounding the NMR sample. The current is then amplified and measured as voltage per unit time. To correct the frequency difference between different nuclei populations, the absolute resonance frequencies for the nuclei in the sample (*f* (sample) are expressed relative to the absolute resonance frequency of a standard reference compound (*f* (reference) such as tetramethylsilane (TMS) and are converted into a chemical shift (δ) in parts per million (ppm). Although absolute resonance frequencies are strictly dependant on the applied magnetic field, chemical shift values are independent of the external magnetic field strength and are therefore expressed in dimensionless units of ppm which provides a more manageable scale. Conversion of the absolute resonance frequencies of the sample (*f* (sample) and the reference compound (*f* (reference) into a chemical shift (δ) in parts per million (ppm) is achieved by using the following formula:

$$\delta\,(\text{ppm}) = \frac{f(\text{sample}) - f(\text{reference})}{f(\text{reference})} \times 10^6$$

### 1.9.1.2 The origin of the NMR signal

In order to provide a further explanation of how an NMR signal arises, it is first necessary to adopt an alternative conceptual framework. In NMR parlance, this conceptual

framework is commonly referred to as the vector model, and provides an important rationale to account for the behaviour of individual nuclei and their associated magnetic moments in the presence of an externally applied magnetic field. The vector model is outlined briefly below.

As mentioned in section 1.9.1.1, individual atomic nuclei can be imagined as miniature bar magnets with each bar magnet having a magnetic moment which, according to the laws of quantum mechanics, can be aligned in any direction. However, when conducting an NMR experiment, rather than observing the magnetic moments of individual nuclei what we actually measure is the net effect of the total magnetic moments resulting from the entire nuclei population of the sample.

When atomic nuclei are placed in an external magnetic field, the magnetic moments associated with each nuclei will precess around the Z-axis. The precession of the net magnetisation around the Z-axis arises due to a slight difference in population between the two nuclei energy states (termed $\alpha$ and $\beta$), with slightly more nuclei populating the lower energy $\alpha$ state. At equilibrium, these magnetic moments can be thought of as aligning in such a way that their contribution forms a net magnetic field along the direction of the applied magnetic field ($B_0$). This notion is often referred to as the bulk magnetisation of the sample, which is commonly represented by a vector quantity known as the magnetisation vector (Figure 1.11). Importantly, at equilibrium, the magnetisation vector (which represents the bulk contribution of the magnetic moments from the entire nuclei population) can be imagined as aligning along the direction of the applied magnetic field (i.e. in the Z direction). At its core, the vector model centres on the disturbance of this

magnetisation vector away from equilibrium and how its direction can be manipulated with respect to the external magnetic field in order to generate an NMR chemical shift.

At equilibrium, the vector model asserts that the net magnetisation vector precesses about the Z-axis at a frequency known as the Larmor frequency. The rate of this precession is proportional to the strength of the external magnetic field and is thus referred to as the Larmor precession. At the Larmor frequency, precession of the magnetisation vector occurs in such a manner that it tilts away from the Z-axis, sweeping out a cone of precession that maintains a constant angle to the applied magnetic field ($B_0$) (Figure 1.12). In order to detect an NMR signal, the processional motion of the bulk magnetisation vector must be disturbed away from equilibrium. In order to achieve this, a very small $90^o$ radiofrequency pulse is applied along the x-dimension (at or near to the Larmor frequency), which results in tipping the magnetisation vector into the XY plane. The precession of the bulk magnetisation vector in the XY plane then induces a current in the coil surrounding the NMR sample. Following the end of the $90^o$ radiofrequency pulse, the bulk magnetisation vector returns back to equilibrium (from precession in the XY plane to precession around the Z-axis) with the current gradually decaying with time through Transverse relaxation ($T_2$). This decrease in current induction is the NMR signal and is known as the Free Induction Decay (FID) (Figure 1.13). A Fourier transformation is then applied in order to convert the FID signal from the time domain to the frequency domain (Figure 1.14).

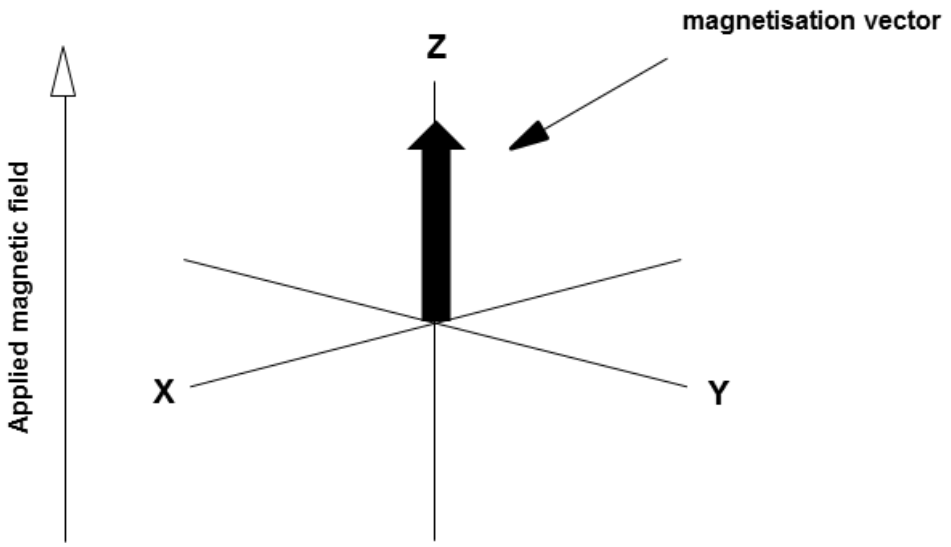



**Figure 1.11 The vector model**. At equilibrium, a sample containing a population of nuclei has a net magnetisation oriented along the Z-axis in the direction of an externally applied magnetic field. In NMR parlance, this is often referred to as the bulk magnetisation vector.

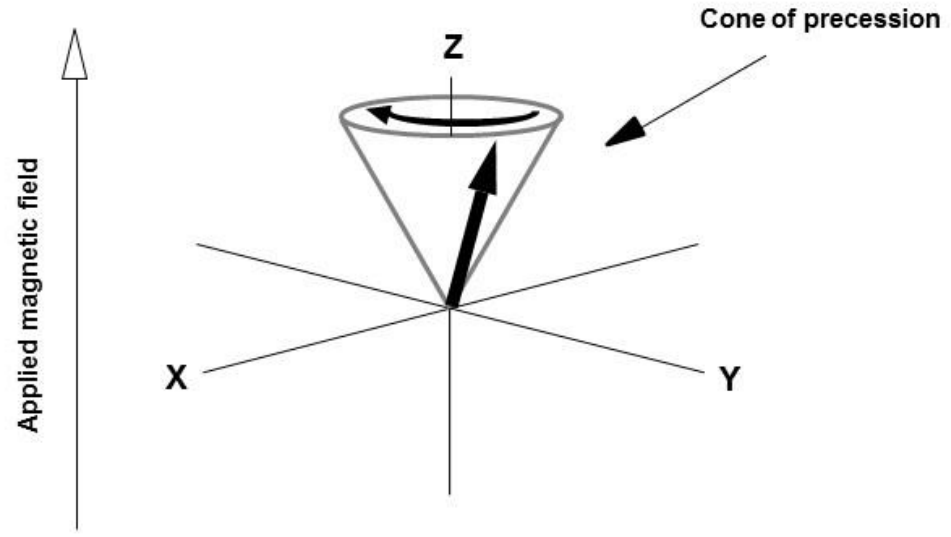
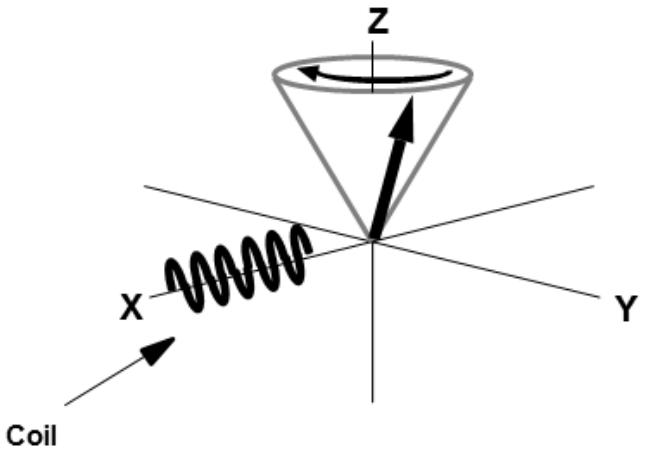



**Figure 1.12 The Larmor precession**. At the particular frequency, the bulk magnetisation vector tilts away from the Z-axis where the vector sweeps out a cone of precession at a constant angle to the applied magnetic field. The rate of precessional motion of the bulk magnetisation vector about the Z-axis is directly proportional to the strength of the applied magnetic field and is referred to as the Larmor precession.




**1.13 Generation of an NMR signal**. Following the application of a $90^{o}$ radiofrequency pulse (at the Larmor frequency) along the X dimension, the precessing magnetisation vector is tipped into the XY plane. Tipping of the bulk magnetisation vector into the XY plane cuts through the coil and induces a current which is then amplified and recorded. For clarity, the coil has only been shown on one side of the x-axis.

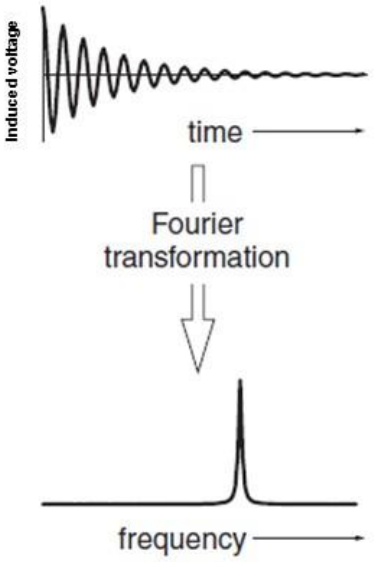



**Figure 1.14 Fourier transformation of the Free Induction decay.** Tipping of the bulk magnetisation vector into the XY plane cuts through the detection coil and induces a current which is then amplified and recorded as an oscillating signal. This oscillating signal is known as the free induction decay and is a phenomenon detected in all pulsed NMR experiments. A chemical shift (measured in frequency) is obtained by applying a Fourier Transformation which converts the time domain signal of the free induction decay into a frequency signal which yields the resulting frequency spectrum of the NMR signal. Figure adapted from Keeler (Understanding NMR spectroscopy).

#### 1.9.1.3 1D $^{1}$H-NMR

The most basic experimental spectrum obtainable on an NMR spectrometer is a 1D $^{1}$H-NMR spectrum. This spectrum provides key information regarding the chemical shift resonances for the different proton environments in a particular molecule. These different proton environments are then utilised for elucidating the structure of small molecule chemicals and often provide a reliable method for verifying the identity of different chemical species when used in conjunction with complementary spectroscopic methods (such as infrared spectroscopy). In addition to small molecule analysis, the versatility of the basic 1D $^{1}$H-NMR experiment can also be extended to larger molecular systems, such as proteins, with a plethora of useful applications. For example, obtaining a 1D $^{1}$H-NMR spectrum enables rapid and reliable determination of whether a particular protein occupies a folded conformation by monitoring the dispersion of proton signals sweeping across the 1-10 ppm range and for a small cluster of high field methyl groups at < 0.5 ppm indicative of a hydrophobic core of a protein. Prior to data collection, water suppression is often conducted to reduce the contribution of the water signal to the resulting NMR spectrum (at a concentration of 55 M, the water signals would dominate the entire NMR spectrum if not suppressed). A range of strategies have been developed that can be employed to achieve solvent suppression yielding good quality spectra [186].

#### 1.9.1.4 2D $^{15}$N-$^{1}$H Heteronuclear Single Quantum Coherence Experiment (HSQC)

The $^{15}$N-$^{1}$H HSQC is a prototypical 2D NMR experiment and is commonly the first two-dimensional NMR experiment conducted on proteins. Whereas 1D $^{1}$H-NMR primarily concerns the detection of one NMR signal (in this case that of $^{1}$H), the basic output of a 2D NMR experiment is a contour plot correlating two NMR signals, either through bond or through space interactions. As such, the $^{15}$N-$^{1}$H HSQC is a classical 2D NMR experiment

correlating $^{15}N$ and $^{1}H$ nuclei, and therefore should detect all NH groups present in a particular molecular species.

Briefly, the $^{15}N$-$^{1}H$ HSQC experiment relies chiefly on the use of a sensitivity enhancement method known as an Insensitive Nuclei Enhanced by Polarisation Transfer (INEPT), a particular radiofrequency pulse sequence which pairs sensitive nuclei (such as $^{1}H$) with insensitive nuclei (such as $^{13}C$ or $^{15}N$) enabling the transfer of magnetisation from the sensitive nuclei to the insensitive nuclei via a mechanism known as scalar coupling.

For large macromolecular systems such as proteins, the $^{15}N$-$^{1}H$ HSQC experiment is an inherently information-rich method and provides multiple layers of information for characterising the overall folds and dynamics of protein structure in solution. The output of a $^{15}N$-$^{1}H$ HSQC experiment is a spectrum derived from collecting data in two dimensions: $^{1}H$ nuclei are collected in the direct dimension and the $^{15}N$ nuclei are collected in the indirect dimension. The spectral resolution of the data collected in each of these dimensions during a $^{15}N$-$^{1}H$ HSQC experiment is highly dependent on the data acquisition time, with longer acquisition times generally leading to better resolution of the data in the dimension being detected. In a similar fashion, the relaxation time during a $^{15}N$-$^{1}H$ HSQC experiment provides short delay periods to allow the system to return to a steady state level of magnetisation. The more readily the system is able to return to steady state magnetisation, the greater the magnitude of magnetisation detected.

Each individual peak represented in $^{15}N$-$^{1}H$ HSQC spectra corresponds to the backbone amide group of individual amino acids in the protein, and also includes the side chain amide groups present in Asparagine (Asn), Glutamine (Gln) and Tryptophan (Trp) residues. Proline residues are refractory to detection in $^{15}N$-$^{1}H$ HSQC spectra due to the

absence of an amide proton. In NMR parlance, a $^{15}$N-$^{1}$H HSQC experiment is often referred to as the "fingerprint" of a protein largely due to the unique observations that can be deduced from analysis of the resulting spectra. Firstly, one can assess whether a particular protein is folded by observing the relative dispersion of individual backbone amide signals. Typically, a uniform dispersion of amide signals across the spectrum is often a reliable indicator of a folded protein structure. In particular, proteins dominated by multiple secondary structure elements including helical regions and pleated sheets often display more evenly dispersed cross-peaks.

In contrast to capturing more ordered protein states, $^{15}$N-$^{1}$H HSQC spectra is also sufficiently sensitive for the detection of local regions of structural disorder and globally unfolded protein states. In particular, sharp and poorly dispersed amide peaks emerging in the 8.0-8.5 ppm region of the spectrum strongly correlate with amide protons present within disordered loop regions of protein structure that are largely exposed to solvent. In addition to the number and dispersion of amide peaks, the relative intensities of each of these peaks also conceal information pertaining to the dynamic properties of a particular protein. These properties typically comprise conformational equilibria, interconversions between monomer and multimer transitions and the aggregation state of a protein.

#### 1.9.1.5 Selective Optimised Flip-Angle Short-Transient-HMQC (SOFAST-HMQC)

During the past decade, much effort has been invested into evolving the methodology to enable fast data acquisition for multidimensional NMR experiments. The SOFAST-HMQC method was developed precisely to speed up data acquisition time for the collection of $^{15}$N, $^{1}$H or $^{13}$C, $^{1}$H HSQC spectra on commercial NMR spectrometers whilst preserving both spectral quality and sensitivity [187]. Briefly, the SOFAST-HMQC involves

applying a polychromatic pulse (PC9) to selectively excite amide protons leading to an overall increase in the rate of $^{1}$H-spin lattice relaxation, allowing for very short recycle delays [188].

#### 1.9.1.6 Transverse Relaxation Optimised Spectroscopy (TROSY) HSQC

A common technical problem when collecting NMR HSQC spectra on larger protein systems (i.e. proteins exceeding >20 KDa) concerns the observation of spectral overlap often accompanied with a reduction in spectral sensitivity. These phenomena arise principally as a result of the sheer abundance of amide proton signals and the fast relaxation properties exhibited by larger molecular species (leading to broadened linewidths). In order to overcome these technical difficulties, one has the option of running a TROSY version of a HSQC which extends the upper size limit of protein NMR to protein systems of < 40 KDa. In a HSQC experiment where decoupling has not been applied, amide protons appear as multiplets as a result of J-coupling. These different multiplet components each have different linewidths which arise as a result of the differential interactions between different relaxation mechanisms. For larger protein systems studied at higher magnetic field strengths, the relaxation properties are dominated chiefly by dipole-dipole and chemical shift anisotropy relaxation pathways. In a TROSY experiment, one exploits the multiplet exhibiting the slowest relaxation and therefore observes the one with the sharpest peak [189].

#### 1.9.1.7 3D $^{1}$H-$^{15}$N-$^{13}$C NMR for unambiguous backbone assignment

As previously mentioned, the $^{15}$N-$^{1}$H HSQC experiment is often the first heteronuclear experiment performed to probe protein structure in the solution state. However, in order to facilitate more complex analysis of protein structure, for example, by conducting assignments to map each of the cross-peaks in the $^{15}$N-$^{1}$H HSQC spectra to a particular

amino acid in the protein, the $^{15}N$-$^{1}H$ HSQC spectrum must first be optimised in order to assess the feasibility of a protein for such analysis. Optimisation of $^{15}N$-$^{1}H$ HSQC spectra is typically achieved by adjusting various experimental parameters (e.g. acquisition time, protein concentration and buffer conditions) in order to improve both spectral quality and resolution. This enhancement in spectral quality and resolution is often signified by a uniform dispersion of backbone amide signals which is indicative of a stable and appropriately folded protein. Once an optimised $^{15}N$-$^{1}H$ HSQC spectrum has been obtained, this then serves as a foundation for more complex structural analysis in the form of 3D NMR experiments to permit unambiguous backbone assignment.

A range of experiments are available to facilitate unambiguous assignment of backbone amide peaks in $^{15}N$-$^{1}H$ spectra, each of which provide additional information regarding the alpha carbon ($C_\alpha$), beta carbon ($C_\beta$) as well as the carbonyl (CO) chemical shifts. However, these experiments require additional $^{13}C$-labelling enabling data to be collected in a third dimension. The most commonly employed experiments to facilitate the collection of backbone assignments, the naming of which is ordered according to magnetisation that is transferred sequentially from one nuclei to the other starting from the backbone amide proton (HN) are the HNCO, HNCACO, HNCA, HNCOCA, HNCACB and HNCOCACB. These experiments will be discussed in further detail below.

#### 1.9.1.7.1 HNCO and HN(CA)CO

The HNCO experiment is widely known as the most sensitive of all the experiments used to perform backbone assignments and is often the first 3D experiment performed when deciphering the amenability of a protein for backbone assignment studies [190]. The high sensitivity of this experiment arises principally as a result of the large J-coupling constant

between the carbonyl CO (i-1) residue and the N (i) residue. During the HNCO experiment, magnetisation transfer proceeds from the $^{1}H$ to $^{15}N$ and then selectively to the carbonyl $^{13}C$ via the $^{15}N^{H}$-$^{13}CO$ J-coupling constant. The magnetisation is then transferred back via the $^{15}N$ for detection on $^{1}H$. The general output of the HNCO experiment therefore consists of chemical shift spectra for the (i-1) carbonyl residue (CO) preceding the HN on which the magnetisation is finally detected (Figure 1.15A).

The HN(CA)CO is a complementary experiment often conducted in parallel with the HNCO. Much like the HNCO experiment, the HN(CA)CO also yields chemical shifts of the preceding (i-1) carbonyl residue (CO) connected to the originally magnetised HN but also provides additional chemical shift information for the CO (i) residue (Figure 1.15B). The HN(CA)CO therefore gives two CO peaks of different intensity, with the stronger peak corresponding to the CO (i) and the weaker peak corresponding to the CO (i-1) residue. Taken together, both the HNCO and HN(CA)CO experiments provide corroborative information regarding the chemical shift of the preceding (i-1) carbonyl residue (CO) connected to the HN from which the magnetisation was initially transferred and thus provides a method for distinguishing between the carbonyl CO (i) and CO (i-1) for each NH group present in a particular protein. However, in comparison to the HNCO experiment the sensitivity of the chemical shift data collected from the HN(CA)CO experiment is significantly reduced owing to comparatively smaller J-coupling constant between the $^{15}N^{H}$-$^{13}C\alpha$ pair. These complementary insights from both the HNCO and HN(CA)CO experiments can therefore be used for the sequential assignment of backbone amide groups present in a particular protein under investigation.

#### 1.9.1.7.2 HNCA and HN(CO)CA

In a similar manner to obtaining chemical shift data for the carbonyl CO of the peptide backbone via HNCO and HN(CA)CO experiments, HNCA and HN(CO)CA experiments can be performed in order to yield chemical shift information between the amide HN and the $^{13}C\alpha$ of both the (i) and preceding (i-1) residues of the peptide backbone. The HNCA experiment yields chemical shift resonances for the coupling between HN and $^{13}C\alpha$ of both (i) and the preceding (i-1) residues of the peptide backbone [191]. However, the HN(CO)CA experiment provides chemical shift resonance information for the coupling between the $^{15}N^{H}$-$^{13}C\alpha$ pair of the preceding (i-1) residue only. Superimposing the spectra from HNCA and HN(CO)CA experiments enables unambiguous assignment of the $^{13}C\alpha$ for both (i) and the preceding (i-1) residues of the peptide backbone (Figures 1.15C and 1.15D).

#### 1.9.1.7.3 HNCACB and HN(CO)CACB

The HNCACB experiment provides chemical shift information based on the coupling between HN and the Cα and Cβ of (i) and the preceding (i-1) residues of the peptide backbone. In contrast, the HN(CO)CACB experiment generates chemical shift information between the (i-1) of Cβ [192]. Superimposing HN(CO)CACB and HNCACB data provides a means of unambiguously identifying the (i-1) Cβ peaks within the HNCACB spectra. Although both HNCACB and HN(CO)CACB experiments are inherently information rich, they are insufficient for unambiguous assignment of peptide backbones. In order to accomplish unambiguous peptide backbone assignment, all of the assignment experiments described must be analysed together in order to increase the likelihood of establishing a correct assignment (Figures 1.15E and 1.15F)

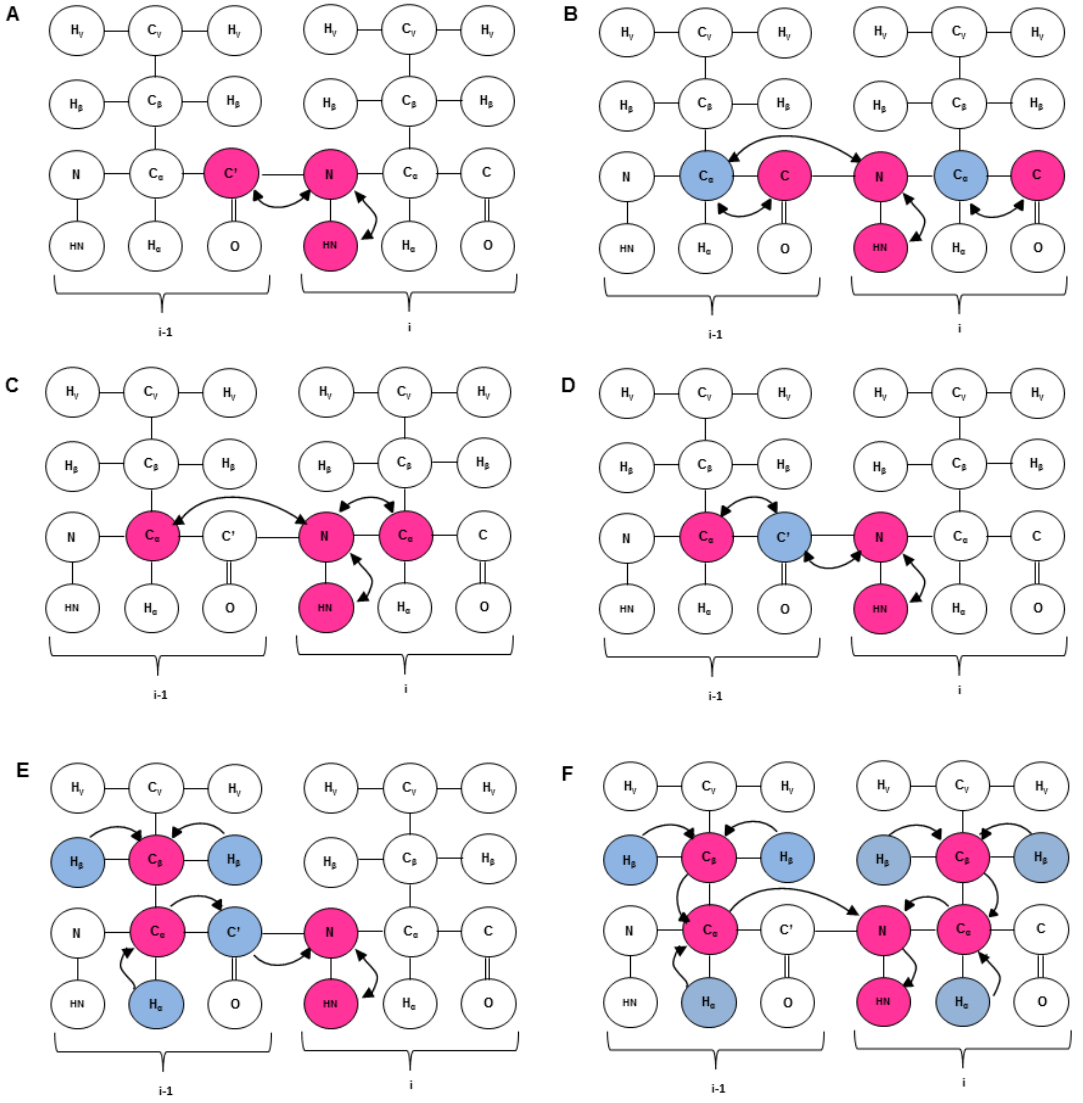



**Figure 1.15 Schematic depiction of 3D backbone assignment magnetisation transfer.** Representation of magnetisation transfer pathways obtained from classical 3D protein backbone assignment spectra. In each experiment, two residues are depicted with individual atoms denoted in circles. To emulate classic NMR nomenclature, residues are denoted as (i) with the preceding residue to (i) denoted as (i-1). Pink circles represent atoms where magnetisation is transferred and data is collected indirectly. Blue circles represent atoms where magnetisation is transferred but no data is collected. Magnetisation transfer experiments for backbone assignments include: **(A)** HNCO **(B)** HN(CA)CO **(C)** HNCA **(D)** HN(CO)CA **(E)** HN(CO)CACB and **(F)** HNCACB. For experiments (A-D) the magnetisation begins on atom HN, but for experiments (E-F) the magnetisation begins at the Hα and Hβ for (E-F). Figure has been adapted from protein-nmr.org.uk, Higman V.

#### 1.9.1.8 Saturation transfer difference NMR

As well as enabling the characterisation of protein structure in the solution state, the versatility of NMR as an analytical tool can also be applied to probe small molecule ligand binding to macromolecular protein structures. Typically, the utility of NMR for the interrogation of protein-ligand interactions can be broadly divided into two different experimental approaches: protein-observed and ligand-observed methods. In each of these experiments, perturbation of the NMR signals of the protein (in the case of protein-observed experiments) or the small molecule ligand (in the case of ligand-observed experiments) are selectively detected to compare the free and bound states of each molecule. In the latter case, one of the most important ligand-observed NMR experiments is Saturation Transfer Difference NMR (STD NMR) which relies chiefly on the exchange between bound and free states of weakly binding small molecule ligands and a larger macromolecule (typically protein).

The rationale underpinning the STD NMR experiment is based on the notion of saturation transfer occurring between protein-ligand interactions in solution. Due to their larger molecular size, protein-ligand complexes possess a slower tumbling rate in solution resulting in a faster rate of relaxation. Conversely, free ligands possess a faster tumbling rate in solution and a correspondingly slower rate of relaxation. During the formation of a protein-ligand complex, the association of the newly bound ligand (as part of the protein-ligand complex) increases the ligand relaxation rate relative to that of the free ligand. Consequently, the rate of ligand dissociation must be faster than the rate of relaxation of the protein-ligand complex otherwise relaxation occurs and the magnetisation transfer would be lost. In an STD experiment, saturation transfer from protein to ligand only occurs under conditions where the exchange rate between the free and bound ligand is fast enough

to accumulate a population of saturated (magnetised) ligand in solution. Importantly, the rate of dissociation of the protein-ligand complex must be large enough to generate sufficiently high concentrations of free ligand in solution, but not so large that it does not permit the optimal ligand residence time necessary to allow efficient saturation transfer from the protein to the ligand to occur [193].

The execution of a typical STD experiment involves a number of steps. The first step involves recording an on-resonance spectrum in which the protein is selectively saturated by irradiating the sample with a radiofrequency pulse that contains only resonances of the protein (e.g. 0 ppm – 1 ppm). The second objective involves recording an off-resonance experiment where a reference spectrum is recorded under the same conditions as the on-resonance spectrum, but instead with the selective radiofrequency pulse off-resonance such that the frequency is set far away from any resonances present in the system. The off-resonance experiment thus yields a spectrum identical to the on-resonance spectrum except that the resonances intensities are not affected by saturation. By subtracting the on-resonance spectra from the off-resonance spectra, only the ligand resonances that experienced a saturation transfer from the protein will be observable in the difference spectra. The magnitude of the resulting saturation difference signals will therefore be inextricably linked to both the efficiency of the protein-ligand saturation transfer (which itself is dependent on the proximity of the ligand to the protein) and the number of protein-bound ligands that received saturation transfer. Non-binding ligands will receive no saturation transfer from the protein and thus their signals will be of an equal intensity in both the on-resonance and off-resonance spectra resulting in no ligand signals appearing in the final difference spectra [193] (Figure 1.16).

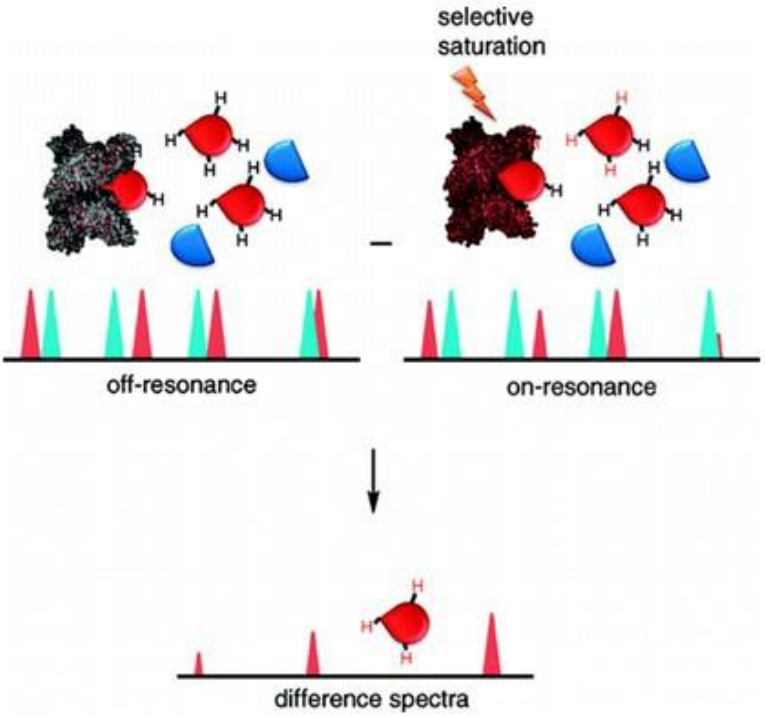



**Figure 1.16 Saturation transfer difference NMR for the detection of ligand binding**. The STD NMR experiment operates by exploiting the difference between bound and free states of weakly binding small molecule ligands to larger molecules, such as proteins. An on-resonance spectrum is first recorded in which the protein is selectively saturated via irradiation of the sample with a radiofrequency pulse that contains only resonances of the protein (e.g. 0 ppm – 1 ppm). An off-resonance spectrum is then recorded under the same conditions as the on-resonance spectrum, but with the selective radiofrequency pulse off-resonance such that the frequency is set far away from any resonances present in the system. The off-resonance experiment thus yields a spectrum identical to the on-resonance spectrum except that the resonance intensities are unaffected by saturation. Subtraction of the on-resonance spectra from the off-resonance spectra generates a difference spectrum where only the ligand resonances that experienced a saturation transfer from the protein will be observable. Figure taken from [194].

#### 1.9.1.9 waterLOGSY

Another common ligand-observed NMR experiment often employed for the detection of protein-ligand interactions is known as the waterLOGSY experiment. Like STD NMR, waterLOGSY involves the transfer of magnetisation via an intermolecular NOE (Nuclear Overhauser Effect – defined as the change in intensity of a particular resonance when the spin properties of another are disturbed from equilibrium) to the transiently bound ligand. However, in the case of waterLOGSY bulk water is irradiated and the magnetisation is transferred from the transiently bound water to the bound ligand. This transfer of magnetisation from water to the bound ligand can occur via multiple different pathways. These include direct transfer of magnetisation from water molecules occupying the ligand binding site, chemical exchange between irradiated water molecules and labile protons on the protein surface, and transfer of magnetisation from water molecules populating the protein surface via the protein-ligand complex [176, 195]. In all of these mechanisms the ligands transiently interact with water via water-ligand protein or protein-ligand complexes, where in both cases the rotational correlation times are slower (relative to free water which has faster tumbling rate) thus exhibiting a negative NOE value with water. In contrast, small molecule ligands that interact solely with bulk water (i.e. non-binders) will have a faster tumbling rate giving a positive NOE. Thus, in a waterLOGSY spectrum one typically observes opposite signs for free vs protein-bound ligands enabling discrimination between binding ligands vs non-binding ligands (Figure 1.17).

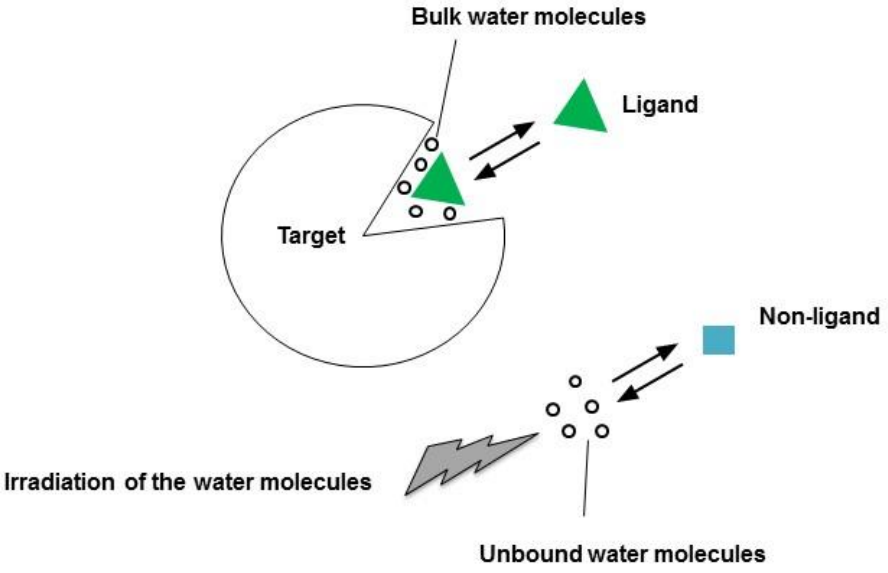



**Figure 1.17 waterLOGSY NMR for the detection of ligand binding**. In a waterLOGSY experiment, magnetization is transferred from the bulk water to the free ligand via the protein-ligand complex via an intermolecular NOE (Nuclear Overhauser Effect – defined as the change in intensity of a particular resonance when the spin properties of another are disturbed from equilibrium) to the transiently bound ligand.

### 1.9.2 Surface plasmon resonance

The characterisation of biomolecular interactions is fundamental for unravelling the molecular basis of cellular processes and is essential for guiding the discovery and development of novel drug molecules for diverse biological targets. Surface plasmon resonance (SPR) has emerged as a powerful optical detection method that provides crucial insights into the binding parameters governing molecular interactions of central biological importance.

The phenomenon of SPR arises when an incident beam of polarised light passing through a medium of high refractive index (such as a glass prism) undergoes total internal reflection by striking the glass slide of a sensor surface. The glass face of the sensor surface is coated with a thin layer of gold that directly interfaces to a buffer solution of low refractive index. In order to generate a resonance signal, the incident beam of polarised light is directed towards the glass face of the sensor surface and the internally reflected light emitted is detected. When the incident polarised light strikes the glass surface interface an evanescent wave is generated which is subsequently absorbed by the electrons present within the gold layer causing them to resonate. The resonating electrons (also referred to as plasmon waves) lead to a reduction in the intensity of the internally reflected light. Consequently, the resonant angle at which the reflected light shows the most significant reduction in intensity is a function of the refractive index of the buffer solution on the opposite side of the sensor surface. As such, the SPR response is exquisitely sensitive to environmental perturbations occurring at the sensor interface facing the buffer solution, and it is precisely this sensitivity of SPR that is exploited for the measurement of biomolecular interactions. By immobilising ligands to the sensor surface, the refractive index of the surface changes

by altering the resonant angle at which the polarised light (of reduced intensity) is reflected. The change in refractive index caused through either the binding or dissociation of a particular analyte to an immobilised ligand is therefore proportional to the mass of analyte bound to the sensor surface. In order to monitor the progression of a binding interaction between an analyte and an immobilised ligand, the SPR instrument generates a sensorgram profile (Figure 1.18) from which important phases of the binding interaction (such as association, kinetics and dissociation) can be derived [196].

As a biophysical technique, SPR benefits from providing a label-free approach which allows for direct measurement and observation of binding in real time, and spares the need to infer binding from indirect measurements through the displacement of a radio- or fluorescently-labelled ligand. In order to measure the kinetics of biomolecular interactions, the biosensor surface is primed for the immobilisation of a particular ligand which can be achieved through a myriad of different coupling approaches. Typical coupling strategies include covalent methods such as amine, thiol or aldehyde coupling, or non-covalent methods such as the employment of metal ion chelation, antibody tags or other affinity tags such as biotin or streptavidin. SPR biosensor assays can be employed to elucidate a number of parameters underlying the formation of molecular complexes. These include information regarding the kinetics (association $k_a$ and dissociation rates $k_d$), affinities at equilibrium ($K_D$) and the stoichiometry of the interaction.

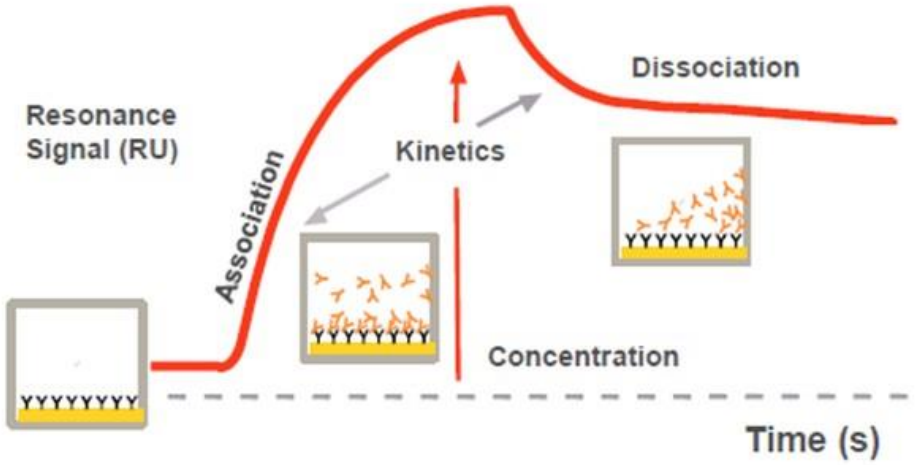



**Figure 1.18. Surface plasmon resonance**. A schematic representation of a typical sensorgram profile for a surface plasmon resonance experiment. The biosensor surface is first primed by covalent or non-covalent immobilisation of a specific ligand. An analyte is then injected over the biosensor surface which leads to the formation of analyte-ligand interactions. This is known as the association phase of the sensorgram and is accompanied by a concomitant increase in the resonance signal. The analyte then dissociates from the immobilised ligand leading to a gradual decay in the resonance signal until it eventually returns back to baseline. The SPR experiment provides a quantitative description of the $k_{on}$ and $k_{off}$ rates of association and dissociation respectively, and enables a dissociation constant $K_D$ to be calculated at the steady state phase of the sensorgram curve.

### 1.9.3 Circular Dichroism Spectroscopy

The quantification of protein secondary structure (i.e. the percentage of α-helices and β-sheets) and the extent of protein folding can be reliably determined by applying the technique of Circular Dichroism Spectroscopy. The technique of circular dichroism involves passing a beam of left and right-handed circularly polarised light at wavelengths between 190-260 nm through a protein sample of interest. During irradiation of the protein sample, the presence of secondary structure elements and chiral molecules leads to the differential absorption of left and right-handed polarised light in different directions. Quantification of the chirality or secondary structure characteristics of the protein can then be determined by the analysis of the resulting spectra.

Typically, circular dichroism experiments involve recording spectra for both the protein sample and the control sample, with the latter containing buffer identical to that which the protein sample has been dialysed into. To ensure the collection of high quality data, all buffers subjected to circular dichroism analysis should be prepared meticulously so that the final buffer contains the fewest possible chemical reagents. In particular, chloride ions ($Cl^-$) should be present either at minimal levels or ideally avoided completely in order to limit quenching of the absorption signal [197].

After data collection, a solvent subtraction step is performed where the buffer control sample is subtracted from the sample containing protein. The resulting spectrum corresponds to an absorption spectrum representing the secondary structure elements of the protein sample. Once processed, the resulting spectra are usually reported as molar ellipticity (θ) in millidegrees (mdeg) as a function of wavelength. The data analysis is usually conducted using the online software package DICHROWEB [198].

### 1.9.4 Analytical Ultracentrifugation

The biophysical technique of analytical ultracentrifugation provides information concerning the homogeneity of both the mass and conformation of particles in solution. In order to achieve this, samples are placed in a sample cell inside an analytical ultracentrifuge and are then subjected to a centrifugal force where the sedimentation velocity of particles in response to the centrifugal force is measured. The sedimentation velocity is inextricably linked to both the molecular mass and the shape (hydrodynamic radius) of the particles in solution. As such, analytical ultracentrifugation is therefore a valuable method for deciphering the homogeneity (in both mass and conformation), aggregation state and oligomerisation of the sample. The concentration distribution of individual particles in the sample is measured across the cell (as a sedimentation gradient) by taking an absorbance reading at a wavelength of 280 nm. The sedimentation velocity measurement is then used to measure the sedimentation co-efficient expressed in the units of Svedbergs (S). Analysis of analytical ultracentrifugation data is typically performed using the software package SEDFIT [199] (see Appendix figure 4). In this work, sedimentation velocity experiments will be conducted in order to measure the concentration of the particles in the cell as a function of time in the centrifugal force

### 1.9.5 Conclusion

Despite once being considered as passive housekeeping enzymes, an evolving body of evidence is beginning to shed new light on the previously underappreciated roles of protein tyrosine phosphatases in human physiology and disease. As such, protein tyrosine phosphatases are now widely considered as a family of signalling enzymes with emerging therapeutic potential.

The cytoplasmic PTP SHP2 plays a central role in the propagation of downstream signalling from multiple growth factor and cytokine receptors, and was the first oncogenic PTP to be discovered. Dysregulation of SHP2 signalling represents a pathological hallmark of developmental disorders such as Noonan and LEOPARD syndrome, and plays a central role in driving the oncogenic phenotypes underlying breast cancer and myeloproliferative disease.

Owing to these roles, the potential of SHP2 as a novel oncology drug target has garnered significant interest over the past decade. Indeed, advances in drug development initiatives such as the implementation of fragment-based screening approaches have challenged conventional dogma by generating potent inhibitors for some of the most challenging classes of biological targets. The overarching aim of this thesis involves exploiting fragment-based screening methods to stimulate the development of SHP2 inhibitors that could serve either as chemical probes for interrogating the basic biology of SHP2, or lead to the development of inhibitors with potential therapeutic significance.

In order to stimulate the discovery of novel SHP2 inhibitors, the first objective of this work involves the expression of $^{15}$N-labelled SHP2 catalytic domain and utilising $^{15}$N-$^{1}$H protein-observed NMR spectroscopy to optimise $^{15}$N-$^{1}$H HSQC spectra of SHP2 catalytic domain in solution. NMR backbone assignment experiments will then be conducted on the SHP2 catalytic domain under the optimised experimental conditions in an attempt to illuminate backbone assignments for amino acid residues in the phosphatase active site. The ligand binding behaviour of the SHP2 catalytic domain in solution will also be investigated by conducting titrations with the SHP2 inhibitor, NSC-87877. Chemical shift

perturbations observed in the presence of NSC-87877 will then be mapped to the backbone assignments to uncover the as yet structurally undisclosed binding mode of the NSC-87877 inhibitor. In addition, as a precursor to conducting fragment-based screening of SHP2, the second objective of this work involves conducting a thorough characterisation of published SHP2 inhibitors using a range of biophysical and biochemical assays in order to validate the biochemical activity and binding affinities of these compounds. SHP2 inhibitors demonstrating consistent activity across all assays will then serve as positively validated tool compounds for use in fragment-based screening. Using the validated SHP2 tool compounds, the third objective will involve conducting fragment-based screening of different SHP2 constructs (including the catalytic domain, tandem SH2 domains and full length SHP2) using both SPR and NMR-based (STD NMR and waterLOGSY) methods in order to identify fragment hits against each domain. Validation of fragment screening hits will then be confirmed by re-testing fragments in SPR dose-response, STD and waterLOGSY experiments. The final and fourth objective of this study will involve developing and optimising a soakable crystal system in order to facilitate structure determination of validated fragment hits bound to the SHP2 catalytic and with the ultimate goal of initiating structure-guided optimisation of these fragments through iterative medicinal chemistry in order to generate more potent chemical inhibitors.

# CHAPTER 2: MATERIALS AND METHODS

## 2.1 Molecular biology

### 2.1.1 SHP2 expression plasmids and construct design

DNA encoding the SHP2 catalytic domain (amino acid residues 237-533) cloned into the pNIC28-Bsa4 vector as a Tobacco Etch Virus (TEV)-cleavable N-terminally $His_{x6}$-tagged fusion protein was a kind gift provided by Dr Alastair Barr and Prof Stefan Knapp at the Structural Genomics Consortium, University of Oxford. For the surface plasmon resonance (SPR) experiments all SHP2 constructs were purchased from Genscript, USA. DNA encoding the SHP2 catalytic domain (residues 237-533), tandem-SH2 domains (residues 4-216) and full-length SHP2 (residues 1-597) was codon-optimised for recombinant expression in *Escherichia Coli* (*E. coli*), chemically synthesised and inserted into a custom-made pET15b vector between restriction sites NdeI-XhoI (Genscript, USA). The DNA encoding each SHP2 construct was additionally modified to contain an N-terminal TEV-cleavable $His_{x6}$ tag (MHHHHHHSSGVDLGTENLYFQS) along with a C-terminal Avi-Tag™ sequence (GLNDIFEAQKIEWHE) to aid biotinylation of each construct following heterologous expression and protein purification. A list of all constructs used in this study is provided in Table 2.1:

| Protein construct | Vector type | Antibiotic resistance | Affinity tag | Cleavage tag |
|---|---|---|---|---|
| **SHP2 catalytic domain** (aa residues 237-533) | pNIC28-Bsa4 | Kanamycin | N-terminal $His_{x6}$ tag | TEV-cleavable |
| **Tandem-SH2 domain** (aa residues 4-216) | pNIC28-Bsa4 | Kanamycin | N-terminal $His_{x6}$ tag | TEV-cleavable |
| **Full length SHP2** (aa residues 1-597) | pET-26b(+) | Kanamycin | C-terminal $His_{x6}$ tag | TEV-cleavable |
| **SHP2 catalytic domain Avi-tag™ (SPR)** (aa residues 237-533) | pET-15b(+) | Kanamycin | N-terminal $His_{x6}$ tag | TEV-cleavable |
| **Tandem-SH2 domain Avi-tag™ (SPR)** (aa residues 4-216) | pET-15b(+) | Kanamycin | N-terminal $His_{x6}$ tag | TEV-cleavable |
| **Full length SHP2 Avi-tag™ (SPR)** (aa residues 1-597) | pET-15b(+) | Kanamycin | N-terminal $His_{x6}$ tag | TEV-cleavable |

**Table 2.1 SHP2 DNA constructs**. The table shows the different SHP2 protein constructs with corresponding amino acid numbers, as well as expression vectors, antibiotic resistance markers, and affinity purification and cleavage tags for each SHP2 construct used in this study. Further information regarding the DNA and amino acid sequences of gene inserts for individual constructs can be found in the Appendix figure 2.

## 2.2 Protein expression

### 2.2.1 Plasmid transformation into DH5α cells

Plasmid DNA (~1 µl) at a concentration of 5-500 ng/µl was used to transform 20 µl of DH5α competent cells (Invitrogen). Competent DH5α cells were then incubated on ice for 20 min at 4$^{o}$C followed by heat shock treatment at 42$^{o}$C for exactly 30 s. Transformed cells were then incubated on ice for a further 2 mins, after which 80 µl of room temperature Super Optimal Broth (SOC) or Lysogeny Broth (LB) was added and the cells left to incubate at 37$^{o}$C with shaking at 210 rpm for 45 mins Transformed cells were plated onto LB agar plates supplemented with the appropriate antibiotic and left to incubate overnight at 37$^{o}$C for colonies to develop.

### 2.2.2 Isolation and amplification of SHP2 DNA constructs

All DNA constructs used in this study were amplified according to the methodology described in the Qiagen Miniprep kit (Qiagen). SHP2 DNA constructs were transformed (see section 2.21 for transformation protocol) into *E. coli* DH5α competent cells, plated and then incubated overnight at 37$^{o}$C. Bacterial colonies were then picked and used to inoculate 5 ml of LB then left to incubate with shaking at 220 rpm at 37$^{o}$C overnight. Overnight bacterial cultures were then centrifuged at 8000 rpm for 3 minutes at 20$^{o}$C in order to pellet cells. After centrifugation, bacterial pellets were thoroughly resuspended by addition of 250 µl P1 resuspension buffer and then lysed by addition of 250 µl of P2 lysis buffer. After 5 min, the lysate mixture was thoroughly mixed with 350 µl of buffer N3 and subsequently centrifuged for 10 minutes at 13,000 rpm in a table-top microfuge. The supernatant was subsequently loaded onto a QIAprep spin column (Qiagen) and centrifuged for 1 min after which the flow through was discarded and the column washed

by addition of 500 µl Buffer PB. After discarding the flowthrough, the column was washed with 750 µl buffer PE and then centrifuged for 1 minute to remove residual wash buffer. Purified DNA was eluted from the column by addition of 50 µl buffer EB (10 mM Tris-HCL, pH 8.5), left to stand for 1 min and then centrifuged for a further 1 min. The concentration of purified DNA was quantified using a Nanodrop and stored at -20$^{o}$C.

### 2.2.3 Large scale expression of unlabelled SHP2 catalytic domain

The pNIC28-Bsa4.SHP2 catalytic recombinant vector DNA (SGC, University of Oxford) was used to transform *E. coli* BL21 (DE3) cells according to the transformation protocol described previously. After overnight incubation at 37$^{o}$C, successful transformants appearing as colonies were then picked and used to inoculate 6 x 10 ml of overnight LB media supplemented with Kanamycin (30 µg/ml). Cultures were then left to incubate overnight at 37$^{o}$C with shaking at 220 rpm. Overnight starter cultures were then used to inoculate 6 L LB media (20 ml overnight culture per 2 L LB media) supplemented with Kanamycin (30 µg/ml) and grown until an $OD_{600}$ reading of 0.5 was reached after which point the temperature was then reduced to 18$^{o}$C. Expression of SHP2 catalytic domain was then induced by addition of 1 mM IPTG when an $OD_{600}$ reading of 0.7 was reached and then left to incubate for 16 hours overnight.

### 2.2.4 Expression of $^{15}$N-1abelled SHP2 catalytic domain for NMR studies

The pNIC28-Bsa4.SHP2 catalytic recombinant vector DNA (SGC, University of Oxford) was used to transform *E. coli* BL21 (DE3) cells according to the transformation protocol (described in 2.2.1). For production of $^{15}$N-labelled SHP2 catalytic domain the expression protocol (described in 2.2.4) was used and suitably modified only by exchanging LB media for M9 minimal media to enable production of $^{15}$N-labelled protein. A standard M9

nutrient mix was added to the M9 minimal media before inoculation with the bacterial starter cultures.

### 2.2.5 Expression of $^{2}H$-$^{15}N$-$^{13}C$ isotopically-labelled SHP2 catalytic domain for backbone assignment determination

For the production of $^{2}H$-$^{15}N$-$^{13}C$ isotopically-labelled SHP2 catalytic domain the same protocols were used for transformation of SHP2 constructs in *E. coli* BL21 (DE3) cells (described in 2.2.1) and expression (described in 2.2.3) and purification (described in 2.3.3 and 2.3.4) of $^{15}N$-labelled SHP2 catalytic domain were employed with slight modifications to enable additional labelling of the protein with $^{2}H$ and $^{13}C$ nuclei. Briefly, $^{2}H$-$^{15}N$-$^{13}C$ isotopically-labelled SHP2 catalytic domain was expressed in 4 L *E. coli* BL21 (*DE3*) cell cultures supplemented with M9 nutrient mix containing 1g/l $^{15}NH_4Cl$, 4 g/l $^{2}H$-$^{13}C$-D glucose, 99% $D_2O$, Kanamycin (30 μg/ml) and grown until an $OD_{600}$ 0.4 was reached after which point the temperature was then reduced to 18$^{o}$C. Expression of $^{2}H$-$^{15}N$-$^{13}C$ isotopically-labelled SHP2 catalytic domain was then induced by addition of 1 mM IPTG when an $OD_{600}$ reading of 0.6 was reached and then left to incubate for 16 hours at 18$^{o}$C overnight.

### 2.2.6 Expression of SHP2 Avi-tag™ SPR constructs

Expression of Avi-tag™ SPR constructs for SHP2 catalytic domain (residues 237-533), tandem-SH2 domains (residues 4-216) and full-length SHP2 (residues 1-597) was facilitated using the same protocols for transformation into DH5α cells (see section 2.2.1), expression (see section 2.2.3) and purification (see section 2.3) of labelled SHP2 catalytic domain.

## 2.3 Protein purification

### 2.3.1 Harvesting cells by centrifugation

After overnight expression at 18$^{o}$C, BL21 (DE3) *E. coli* cells were harvested by centrifugation at 6000 rpm for 20 mins at 4$^{o}$C (Avanti J-20XP centrifuge using a JLA 8.1 rotor). After discarding the supernatant, cell pellets were resuspended in resuspension buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, 50 mM Imidazole, 0.5 mM TCEP (tris(2-carboxyethyl)phosphine) and two EDTA (ethylenediaminetetraacetic acid)-free protease inhibitor cocktail tablets (Roche). Resuspended cells were then stored at -80$^{o}$C for further use.

### 2.3.2 Mechanical homogenisation of cells

Cells pellets were homogenised at 15,000 psi for three continuous cycles using an Emulsiflex C-3 (Avestin). During continuous rounds of lysis cells were incubated at 4$^{o}$C on ice. Fully homogenised crude lysates were centrifuged at 25000 rpm for 45 mins at 4$^{o}$C (Avanti J-25, centrifuge using a JA25.50 rotor). The supernatant was then filtered using a 0.45 μm filter.

### 2.3.3 Purification of His-tagged proteins by $Ni^{2+}$-NTA affinity chromatography

Purification of all unlabelled and isotopically-labelled SHP2 constructs for NMR and SPR studies (all of which were $His_{x6}$-tagged) was conducted using the same method of $Ni^{2+}$-NTA affinity chromatography. After passing the supernatant though a 0.2 μm syringe filter, $His_{x6}$-tagged proteins were loaded onto a 5 ml His-trap column (GE Healthcare) pre-incubated with re-suspension buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, 50 mM Imidazole, 0.5 mM TCEP and two EDTA-free protease inhibitor cocktail tablets

(Roche) and pre-washed according to manufacturer's instructions before being purified by $Ni^{2+}$-NTA affinity chromatography. His-tagged SHP2 protein was eluted using elution buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM Imidazole, 0.5 mM TCEP and two complete EDTA-free protease inhibitor tablets (Roche). Elution fractions were tested for protein by addition of Bradford Reagent and those fractions containing the most protein were run by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) to determine purity and ascertain molecular weight in conjunction with known molecular weight markers (Bio-Rad). For all SHP2 constructs a further purification step was required using size-exclusion chromatography (see 2.3.5).

### 2.3.4 Removal of $His_{x6}$-tag by incubation with TEV protease

Elution fractions containing the highest levels of $His_{x6}$-tagged protein were pooled and transferred to SnakeSkin™ dialysis tubing (Thermo Scientific) before overnight dialysis in 50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP in the presence of TEV protease (2 mg/ml). In all cases, a second His-clean step was employed by passing the overnight dialysate back through a pre-equilibrated $Ni^{2+}$-NTA affinity column in order to remove the cleaved $His_{x6}$-tag and any uncleaved protein. Pre and post-TEV cleavage samples, along with flow-through, wash and elution fractions were subject to SDS-PAGE analysis in order to validate successful cleavage. Successfully cleaved protein was most abundant in the wash fraction, which was then concentrated to ~ 5 ml and further purified by size-exclusion chromatography (see section 2.3.5).

### 2.3.5 Protein purification by size-exclusion chromatography

His-cleaved protein was finally purified to homogeneity using an Superdex-75 26/60 size-exclusion column (GE Healthcare). Flow rates were set at 2.5 ml/min with a 4 ml fraction collection volume. Elution fractions corresponding to the highest absorbance peaks at 280 nm were analysed by SDS-PAGE to gauge purity and fractions containing the highest quantities of pure protein were pooled and concentrated prior quantification of protein concentration.

## 2.4 Protein biochemical analysis

### 2.4.1 Determination of protein purity by SDS-PAGE

Protein samples (~1-20 μl) for SDS-PAGE analysis were prepared by mixing samples with equal volumes of 2x Laemmlli buffer (Sigma) and samples were then loaded onto an 18-well pre-cast 4-12% SDS-PAGE gel (Bio-Rad). Molecular weight markers (Bio-Rad) were run as a reference alongside protein samples for comparison of molecular weight to known standards. Typically, SDS-PAGE gels were run at 180V for 45 mins in 2-(N-morpholino)ethanesulfonic acid (MES) SDS running buffer (Bio-Rad). All SDS gels were stained with Instant-blue$^{TM}$ reagent and then left to develop while shaking for 30 mins.

### 2.4.2 Determination of protein concentration by $A_{280}$ measurement

Following size-exclusion purification, proteins were concentrated using an Amicon Ultra$^{TM}$ filter with a 10 KDa cut-off. Protein samples were then centrifuged at 4000 rpm for 10 min cycles at 4$^{o}$C until reaching an approximate volume of ~1 ml. Determination of the final concentration of purified protein was facilitated by taking absorbance readings at a wavelength of 280 nm using a UV spectrometer (Shimadzu UV-1700 Pharmaspec).

Quantification of the final protein concentration was conducted by computing the absorbance reading at 280 nm and the molar extinction coefficient for the protein sequence (obtained from ProtParam; see Table 2.2) into the Beer-Lambert equation:

$$A = \varepsilon cl$$

Where A = absorbance reading measured at 280 nm,

$\varepsilon$ = molar extinction coefficient ($M^{-1}\,cm^{-1}$)

c = concentration (M)

l = path length (cm)

| Protein construct | Molar extinction coefficient ($M^{-1}$ $cm^{-1}$) |
|---|---|
| **SHP2 catalytic domain** (aa residues 237-533) | 45380 |
| **Tandem SH2 domain** (aa residues 4-216) | 22920 |
| **Full length SHP2** (aa residues 1-597) | 72770 |
| **SHP2 catalytic domain Avi-tag™ (SPR)** (aa residues 237-533) | 50880 |
| **Tandem SH2 domain Avi-tag™ (SPR)** (aa residues 4-216) | 28420 |
| **Full length SHP2 Avi-tag™ (SPR)** (aa residues 1-597) | 78270 |

**Table 2.2 Molar extinction coefficient values for SHP2 constructs.** Molar extinction coefficient values were calculated using the ProtParam software and then inserted into the Beer-Lambert equation along with the $A_{280}$ readings to determine the concentration of each SHP2 construct in solution.

## 2.5 Biophysical characterisation of protein structure

### 2.5.1 Analytical Ultracentrifugation (AUC)

Analytical ultracentrifugation (AUC) experiments were performed in collaboration with Rosemary Parslow (School of Biosciences). Unlabelled SHP2 catalytic domain was purified to homogeneity by $Ni^{2+}$-NTA chromatography followed by size-exclusion chromatography and dialysed into 50 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.5 mM TCEP, and 50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP. Two concentrations of SHP2 catalytic domain were made up with phosphate and HEPES buffer to a final volume of 500 µl so that the absorbance readings at 280 nm were 0.15 and 0.35 to ensure optimum data collection. Protein samples were centrifuged using a Beckman XL-1 analytical ultracentrifuge containing an An-50 Ti Analytical 8-place rotor at speeds of 25000 and 40000 rpm for 20 h at 20$^{o}$C. Data analysis was conducted using the SEDFIT software [200] (see Appendix figure 4).

### 2.5.2 Secondary structure analysis using Circular Dichroism

Samples for Circular Dichroism (CD) analysis were prepared by dialysis of unlabelled SHP2 catalytic domain into separate buffer solutions containing 20 mM sodium phosphate pH 7.5, 0.5 mM TCEP and 20 mM Tris pH 7.5 and 0.5 mM TCEP. The pH of all buffers was adjusted with appropriate equivalents of $H_2SO_4$ in order to reduce the presence of Cl- ions which would otherwise quench the CD signal. Proteins were diluted to a final concentration of 0.1 mg/ml in a final volume of ~300 µl before transfer to a 1 mm cuvette. All measurements were recorded on a JASCO-J810 spectrometer at 25$^{o}$C in the spectral range of 180 nm-240 nm at 0.5 nm intervals with a 1 s scan rate at each data collection point giving a total of 120 scans. Final CD spectra were obtained by subtraction of the

buffer CD spectra from CD spectra recorded with protein. Subsequently analysis of the final CD spectra was achieved using the CDSSTR analysis tool from the DICHROWEB server [201] (see Appendix figure 4) which enabled quantification of the ratio of protein secondary structure elements.

## 2.6 NMR experiments

### 2.6.1 NMR sample preparation

Samples for nuclear magnetic resonance (NMR) spectroscopy experiments were analysed using Agilent 600 MHz, 800 MHz and 900 MHz spectrometers, each of which was equipped with a triple resonance cryoprobe fitted with the appropriate z-axis pulse field gradients. Prior to data acquisition, the magnetic signal was locked on $D_2O$ to calibrate and stabilise the magnetic field. After locking the signal on $D_2O$, the spectrometer was tuned to each nuclei channel and shimmed in order to homogenise the magnetic field prior to recording high resolution NMR spectra. Excitation sculpting was run on all samples to suppress the water signals enabling more accurate resolution and acquisition of the protein signals. Acquisition of the protein signals was then achieved through calculating the $90^o$ pulse, which represents the time-frame (measured in μs) of the radiofrequency pulse that leads to the rotation of the net magnetisation from the Z-axis to the XY plane. All NMR experiments and analysis were conducted in collaboration with Dr Mark Jeeves (Henry-Welcome NMR Building, University of Birmingham).

### 2.6.2 One-dimensional $^1$H-NMR

$^1$H NMR spectra was consistently recorded prior to conducting multidimensional experiments in order to confirm the presence of folded protein at the experimental concentration, and also to verify the chemical identity and gauge the solubility of various

chemical compounds. Before data acquisition, shimming functions were applied followed by excitation sculpting [202] to suppress water signals and allow determination of the 90$^{o}$ pulse. Parameters used for excitation sculpting include a relaxation time of 1s, a spectral width of 18 ppm for $^{1}$H and an acquisition time of 70 ms.

### 2.6.3 $^{15}$N-$^{1}$H NMR HSQC titrations with compound NSC-87877

$^{15}$N-$^{1}$H NMR heteronuclear single quantum coherence (HSQC) titrations with compound NSC-87877 were performed on a Varian INOVA 800 MHz spectrometer fitted with the appropriate z-axis gradients and a 4-channel cryoprobe. Sequential additions of sodium phosphate (pH 7.5) were added to 0.3 mM $^{15}$N-isotopically labelled SHP2 catalytic domain (dialysed into 50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP) to final concentrations of 5 mM, 10 mM and 20 mM in order to gauge improvements in spectral quality. All HSQC spectra were recorded in the presence of a final NMR buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 10% (v/v) $D_2O$, 5% (v/v) DMSO and 20 mM sodium phosphate to improve spectral resolution. Compound NSC-87877 was solubilised to a final concentration of 200 mM in 100% (v/v) $D_6$-DMSO. $^{15}$N-$^{1}$H TROSY HSQC spectra were recorded on $^{15}$N-isotopically labelled SHP2 catalytic domain (and 20 mM sodium phosphate) in the presence and absence of a two-fold equivalent (0.6 mM) of compound NSC-87877. Analysis was conducted using CCPN analysis software (see Appendix figure 4) and spectra in the presence and absence of inhibitor was overlaid to identify residue-specific perturbations and map these to previously collected backbone assignments.

### 2.6.4 Triple resonance experiments for backbone assignment of SHP2 catalytic domain

NMR triple-resonance experiments for backbone assignment of the SHP2 catalytic domain were performed on a Bruker 900 MHz NMR spectrophotometer equipped with a triple resonance cryoprobe and z-axis pulse field gradients. Backbone assignments were collected by running TROSY versions of HNCO, HNCACO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB experiments [203] along with $^{15}$N-$^{1}$H HSQC on 0.5 mM $^{2}$H-$^{15}$N-$^{13}$C-labelled SHP2 catalytic domain in a final NMR buffer containing 50 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.5 mM TCEP and 10% (v/v) $D_2O$. 25% non-uniform sampling was applied during data collection across 3D planes for all assignment experiments. For the $^{15}$N dimension, 64 planes were collected across all assignment experiments. A total of 64 planes were also collected for the carbonyl (CO) and alpha carbon (CA) dimensions corresponding to HNCO, HN(CA)CO, HNCA and HN(CO)CA experiments, as well as 96 planes collected for the alpha carbon (CA) and beta carbon (CB) corresponding to HNCACB and HNCOCACB experiments. Backbone pulse sequences were obtained from VNMRJ Biopack. Triple resonance data was processed with NMRPipe (see Appendix figure 4) and analysed using CCPN (see Appendix figure 4) analysis in collaboration with Dr Mark Jeeves (Henry-Welcome NMR facility, University of Birmingham).

### 2.6.5 STD-NMR validation of NSC-87877 inhibitor binding

All STD-NMR spectra were acquired on a Varian 600 MHz spectrometer with z-axis pulse field gradients and a triple-channel cryogenic probe. Samples containing 100 μM NSC-87877 and 5 μM unlabelled SHP2 catalytic domain were prepared in glass champagne

vials in buffer containing 50 mM d-Tris pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 5% (v/v) DMSO and 20% (v/v) $D_2O$ to a final volume of 50 µl. Data was analysed in Topspin 3.1.

## 2.7 Enzymatic activity assays

### 2.7.1 pNPP activity assay

The phosphatase activity of SHP2 catalytic domain was assessed using the spectophotometric p-nitrophenyl phosphate (pNPP) assay. Initial reactions to determine the $K_m$ of pNPP for SHP2 were conducted at 25$^oC$ in 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP). Enzymatic reactions were initiated with addition of 100 nM SHP2 to assay mixtures containing 0, 1, 2.5, 5, 10, 20, 30, 40, 50 mM pNPP substrate (~$0.5K_m$-$10K_m$) in final reaction volumes of 100 µl. Phosphatase activity for each pNPP concentration was assayed at time points 0, 2, 4, 6, 8, 10, 12 min and was subsequently arrested with 50 µl of 5M NaOH. Activity measurements for each substrate concentration were corrected for non-enzymatic hydrolysis by subtracting the control at time point 0 in the absence of enzyme. The amount of p-nitrophenol product generated from each reaction was measured by absorbance at 405 nm using a Bio-Rad iMark$^{TM}$ microplate reader. Absorbance readings were plotted against pNPP concentration for each time point and initial rates in Abs $min^{-1}$ were converted into µM $min^{-1}$ by using the molar extinction coefficient 18,000 $M^{-1}$ $cm^{-1}$. Michaelis-Menten curves were generated using Sigma plot 12.0 by fitting data to a rectangular hyperbola enabling derivation of enzymatic parameters $K_m$ and $V_{max}$.

### 2.7.2 Inhibitor activity assays

Inhibition of SHP2 phosphatase activity was assessed using the spectrophotometric pNPP assay in the presence of SHP2 inhibitors IIB08 and NSC-87877 and was conducted using the same experimental protocol and assay parameters as described in 2.7.1. Compound IIB08 was tested at concentrations of 0 µM, 5 µM and 10 µM, and compound NSC-87877 at concentrations of 0 µM, 0.5 µM, 1.0 µM and 2.0 µM in final reaction volumes of 100 µl. In addition, inhibitor assays were repeated in the presence of BSA (0.1 mg/ml) and Tween 20 in order to eliminate non-specific and aggregation-driven mechanisms of inhibition. Michaelis Menten curves were generated using Sigmaplot 12.0 by non-linear regression analysis by fitting to a rectangular hyperbola enabling derivation of enzymatic parameters $K_m$ and $V_{max}$ and determination of the mode of inhibition. Values for the inhibitor constant $K_i$ were derived by fitting the inhibitor curves to a rectangular hyperbola modified for competitive and non-competitive inhibition.

## 2.8 SHP2 fragment screening

### 2.8.1 Preparation of fragment library

A 433-compound fragment library of sufficient diversity was compiled from a selection of commercially available chemical fragments (Maybridge and Sigma). Fragments were selected on the basis of physicochemical properties including molecular weight (typically between 110 - 250 g $mol^{-1}$), predicted solubility in water (2 mM), suitability for the proposed screening strategy, chemical tractability, appropriate functionality for synthesis, diversity of chemical scaffolds and general applicability to a wide array of target classes. Compounds were prepared as 200 mM stocks in 100% (v/v) DMSO in a dark moisture-free cabinet. Prior to screening, 1D $^{1}$H-NMR analysis was conducted on all compounds to

confirm structural integrity and verify chemical identity. Mass spectrometry was also orthogonally employed in order to resolve ambiguities in compound validation.

### 2.8.2 Biotinylation of SHP2 Avi-tag™ constructs for SPR

Purified Avi-tagged SHP2 constructs were buffer exchanged into biotinylation reaction buffer consisting of 100 mM Tris pH 7.5, 20 mM NaCl, 5 mM $MgCl_2$. Avi-tagged SHP2 constructs were then incubated for 2 hours at 30$^{o}$C in a reaction mixture containing 25 x protease inhibitor cocktail (PIC; Roche), 100 mM ATP, 100 µM Biotin and 2 mg/ml BirA ligase enzyme. Biotinylated SHP2 constructs were further purified by size exclusion chromatography (Superdex-75 10/300 GL column; GE Healthcare) to remove excess free biotin. Elution fractions from size-exclusion chromatography were concentrated (see 2.4.1) and the final concentration calculated by measuring absorbance at $A_{280}$ (see 2.4.2).

### 2.8.3 Western blotting analysis

Western blots of purified biotinylated SHP2 catalytic domain, tandem SH2 domain and full length SHP2 were conducted in order to determine whether biotinylation of protein constructs was successful. Protein samples (2 µl) were mixed with equal volumes of 2x Laemelli buffer (Sigma) and run by SDS-PAGE (see 2.4.1) alongside control protein CamKID at 180V for 45 mins. The SDS gel was then washed for 5-10 mins with western blot buffer consisting of 1 x Tris-buffered saline and Tween 20 (TBST) to remove residual gel running buffer and was subsequently transferred to a polyinylidene fluoride (PVDF) and nitrocellulose membrane as part of a Trans-Blot® Turbo™ Mini PVDF Transfer Pack

(Bio-Rad). Transfer was conducted using a Trans-Blot® Turbo™ blotter set at 180 V for 7 mins (Bio-Rad) and after successful transfer the nitrocellulose membrane was washed in 3% (w/v) BSA for 1 hour, followed by incubation with streptavidin-horse radish peroxidase (HRP) antibody (1:20,000 fold dilution) for a further 1 hour at room temperature. After antibody incubation, the nitrocellulose membrane was washed with 1 x TBST for 3 x 10 min washes to remove residual antibody and developed with a 50:50 ratio of ECL (enhanced chemiluminescence solution) solution using the ChemiDoc system (Bio-Rad).

### 2.8.4 Fragment screening by SPR

A Biacore T200 instrument (GE Healthcare) was used to conduct all SPR fragment-based screening experiments. Biotinylated SHP2 catalytic domain, tandem SH2 domains and wild-type full length SHP2 proteins were each immobilised on a Series S SA streptavidin-coated sensor chip using standard methodology and reagents (GE Healthcare handbook). SHP2 catalytic domain was immobilised on flow cell 2 with the tandem SH2 domains and wild-type full length SHP2 immobilised to flow cells 3 and 4, respectively. Flow cell 1 was left unmodified and used as a reference control channel. SHP2 catalytic domain, tandem SH2 domains and wild type full length SHP2 were prepared as 1 μM, 1.5 μM, and 3.0 μM respectively in buffer containing 10 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP and then immobilised to a Series S SA streptavidin-coated sensor chip over a 10 min period with a set immobilisation level of 10000 RU. Final immobilisation levels were 7900 RU, 7875 RU and 7600 RU, for the SHP2 catalytic, tandem SH2 domains and wild-type full length SHP2, respectively. A 433-compound Maybridge/Sigma library was then screened to identify active fragments binding to each SHP2 construct. Prior to running the

fragment screen all compounds were solubilised in 100% (v/v) $D_6$-DMSO to a final concentration of 200 mM. Compound stocks at 200 mM were initially diluted to 40 mM in 100% (v/v) $D_6$-DMSO before further dilution to a final screening concentration of 200 μM. Fragment screening experiments were performed at 25$^{o}$C with assay running buffers consisting of phosphate and non-phosphate buffer conditions. For non-phosphate buffer conditions, fragments were prepared in 10 mM HEPES pH 7.5, 150 mM NaCl, 0.05% (v/v) Tween 20 (Sigma) and 0.5 mM TCEP. In phosphate buffer, fragments were prepared in 1.05 x PBS (GE Healthcare) pH 7.5, 0.05% (v/v) Tween 20 (Sigma) and 0.5 mM TCEP.

For each screening run, fragments were injected over the sensor chip surface at a final concentration of 200 μM with a contact time of 30s and a flow rate of 30 μl/min. The SHP2 catalytic domain inhibitor NSC-87877 at an initial stock concentration of 200 mM in 100% (v/v) $D_6$-DMSO was then subsequently diluted using both phosphate and HEPES running buffers to a running concentration of 100 μM. Compound NSC-87877 was then injected every 46 cycles and run as a positive control throughout the screening assay to gauge protein stability as a measurement of ligand binding throughout the duration of the assay. Before running the fragment screens, a DMSO (dimethyl sulfoxide) solvent correction curve was prepared with either phosphate or 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) fragment dilution buffer spanning a DMSO concentration range between 4.5-5.8% (v/v), corresponding to concentrations above and below the 5% (v/v) final DMSO concentration used in the screening assay. Biacore 2000 evaluation software was then used for the analysis of each screening dataset. Screening data for all experiments was corrected for DMSO responses by applying the solvent correction to all samples in the screen to analyse the extent to which the DMSO

concentration of the fragments matched the DMSO concentration in the running buffer. Observable discrepancies or DMSO mismatches between samples and running buffers above or below the solvent correction parameters were eliminated from the screen and re-tested. Screening data was then normalised according to compound molecular weights and then corrected for signals from the running buffer. Final responses for fragment hits were generated by subtraction of the signals generated from protein-immobilised flow-cells from signals obtained from the unmodified reference surface.

Fragment hits were then selected for by interrogation and inspection of final sensorgram profiles (subtracted from the blank reference cell). Responses showing a classical on/off "table-top" binding profile that gave responses at or greater than the 3 RU cut-off were validated as hits. Validated hits from the primary screen were then tested by dose-response SPR experiments in order to derive appropriate binding affinities.

Fragments hits recorded against the SHP2 catalytic domain by SPR screening were selected for further analysis in dose-response experiments by SPR. Fragments active in dose-response experiments were repurchased as singletons (Maybridge) and prepared as 200 mM stock solutions in 100% (v/v) $D_6$-DMSO. Singleton stock solutions were then prepared as 10 mM stock solutions in 5% (v/v) DMSO by dilution in buffer containing 10 mM HEPES pH 7.5, 150 mM NaCl, 0.05% (v/v) Tween 20 and 0.5 mM TCEP. Dose-response experiments were then conducted in running buffer containing 10 mM HEPES pH 7.5, 150 mM NaCl, 0.05% (v/v) Tween 20, 0.5 mM TCEP and 5% (v/v) DMSO. Biotinylated SHP2 catalytic and full length SHP2 were immobilised to saturation yielding final responses of 7900 RU and 6870 RU, respectively, using the same concentrations and immobilisation parameters as previously described. Fragments were injected with a total

contact time of 30 s at a concentration range of 0.64 mM - 10 mM (1:5 serial dilution series from 10 mM starting concentration) using a flow rate of 30 µl/min. The positive control compound NSC-87877 (Merck) was also tested using a 30 s contact time and 30 µl/min flow rate at 0.0064 µM-100 µM (1:5 serial dilution series from 100 µM) in order to confirm protein stability and assess the reproducibility of the results. Biacore 2000 evaluation software was used to confirm fragments showing dose-response binding and validate fragment binding affinities. Dose-response binding data for each fragment dilution series was corrected for DMSO mismatch and apparent fragment affinities were determined by fitting to a 1:1 protein-ligand binding stoichiometry. The maximum binding capacity of the surface ($R_{max}$) based on the level of protein immobilisation was calculated for each fragment hit and the sensorgrams interrogated for binding profiles to discriminate between specific and non-specific binding.

### 2.8.5 Fragment screening by STD-NMR and waterLOGSY

Fragment screening by STD (saturation transfer difference)-NMR and water-ligand observed via gradient spectroscopy (waterLOGSY) was conducted in collaboration with Dr Mark Jeeves (Henry-Welcome NMR building, University of Birmingham) using the same fragment library used for SPR screening. Out of the 433 fragments only 384 were screened with 49 fragments discarded due to solubility issues. The 384 fragments were screened in cocktails of 4 and each mixture added to 5 µM of unlabelled SHP2 catalytic domain, tandem SH2 domains and full length SHP2. Screening of the fragment cocktails was conducted using STD-NMR and waterLOGSY using two separate screening conditions for each measurement in final NMR buffers containing 50 mM d-Tris pH 7.5,

150 mM NaCl, 0.5 mM TCEP, 20% $D_2O$, 5% (v/v) $D_6$-DMSO and 20 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 20% (v/v) $D_2O$, 5% (v/v) $D_6$-DMSO to compare fragment hit rates in phosphate and HEPES buffer conditions. The final concentration of each fragment in the cocktails was 500 μM. Samples were mixed in champagne vials and subsequently transferred into 1.7 mm NMR tubes using a Gilson GX-271 liquid handler. STD-NMR and waterLOGSY measurements were acquired on a Bruker 600 MHz spectrometer at 298 K equipped with a cryoprobe, the appropriate z-axis pulse field gradients and an autosampler. For STD data acquisition, irradiation frequency parameters were - 0.5 ppm for on resonance spectra and 25 ppm for off resonance spectra, with the power set to 5 mw [193]. The acquisition time used for both STD and waterLOGSY experiments was 1.7s. For STD experiments a saturation time of 3s was used, whereas a mixing time of 1.5s was used for all waterLOGSY experiments. Final protein concentrations of 5 μM and ligand concentrations of 500 μM were used throughout both STD and waterLOGSY experiments. The number of scans used in both STD [193] and waterLOGSY [176] experiments was 128. STD and waterLOGSY data were both analysed using TOPSPIN version 3.2 by comparing experimental STD spectra for each fragment cocktail with $^1$H-NMR spectra for the individual compounds. Fragments showing an inversion of the waterLOGSY signal were validated as hits. Fragment hits in STD spectra were deciphered and subsequently deconvoluted in sample mixtures showing positive STD resonances.

### 2.8.7 Thermal shift assays for buffer stability

Thermofluor$^{TM}$ assays were conducted to determine optimal buffer conditions of the SHP2 catalytic domain as a precursor to crystal screening. A 2 μl aliquot of SYPRO orange

fluorescent dye (20x stock) was added to the protein (20 μM) and then a further 2 μl of the protein:dye mix was added to a 96-well buffer screen (48 μl in each well). Thermal stability measurements for the protein in each buffer condition were obtained by recording the assay in an RT-Mx3005P PCR machine (Stratagene). Melting curves were recorded at temperatures between 25$^{o}$C-90$^{o}$C and fluorescence readings were obtained at excitation and emission wavelengths of 490 nm 560 nm, respectively.

## 2.9 X-ray crystallography

### 2.9.1 Crystallisation screening

Prior to setting up crystallisation trials, SHP2 catalytic domain was expressed and purified to >95% homogeneity as determined by SDS-PAGE analysis according to methods previously described in 2.2.1, 2.3.1 and 2.4.1. Small-scale crystallisation trials were initially set up using a Mosquito nanolitre crystallisation robot (TTP Labtech). This involved mixing 100 nl of protein with an equivalent 100 nl volume of crystallisation buffer in a series of 96-well plates. Conditions yielding promising crystal hits were then progressed and optimised on a larger scale by adjusting precipitant concentration, protein concentration, the ratio of protein to crystallisation buffer in the drop, and the total drop volume. Large scale optimisation using 24-well plates (Cloverleaf) led to growth of diffraction-quality crystals of the SHP2 catalytic domain in conditions containing 8%-9% PEG 3350, 25 mM sodium malate pH 7.5 with 3 μl crystal drops containing 1 μl protein to 2 μl crystallisation buffer. All crystallisation trials were conducted at 5$^{o}$C with crystal growth occurring after 3 days.

### 2.9.2 X-ray data collection and processing

Prior to X-ray data collection, crystals grown at 8 mg/ml protein concentration in crystallisation buffer containing 8%-9% PEG 3350 and 25 mM sodium malate pH 7.5 were soaked in a cryoprotectant mixture containing 25% ethylene glycol before being immediately flash frozen in liquid nitrogen. Crystals flash frozen in liquid nitrogen were then mounted and cooled at 100 K under a nitrogen gas cryostream. X-ray diffraction data was collected to a resolution of 3.1 Å on an in-house MicroMax 007HF X-ray source (Rigaku) using a Saturn CCD detector (X-ray facility, School of Biosciences, University of Birmingham) and processed by Dr Fiyaz Mohammed (Institute of Cancer and Genomic Sciences, University of Birmingham). The initial diffraction data was then integrated, scaled and merged using programs of the XDS suite. Molecular replacement was used to obtain missing phase information using the PHENIX program suite [204] (see Appendix figure 4).

# CHAPTER 3: STRUCTURAL ANALYSIS OF SHP2 INHIBITOR BINDING BY NMR SPECTROSCOPY

## 3.1 Experimental aims

To assist the development of more efficacious inhibitors and enable further understanding of the ligand binding behaviour of the SHP2 catalytic domain in the solution state, the aim of this chapter was to utilise $^{15}N$-$^{1}H$ protein-observed NMR spectroscopy to provide the first structural evidence for the binding mode of the previously identified SHP2 inhibitor, NSC-87877. To do this it was necessary to perform $^{1}H$-$^{15}N$-$^{13}C$ amide backbone assignments. More broadly, utilising an NMR-based approach the aim was to explore ligand-binding to the SHP2 catalytic domain under physiologically relevant conditions and use these insights to determine the optimal conditions for conferring protein stability as a precursor to fragment-based compound library screening and inhibitor development.

The work described in this chapter has built upon previous and unpublished data generated by Dr Timothy Knowles (School of Biosciences, University of Birmingham) and Dr Mark Jeeves (Henry Welcome-NMR Building, University of Birmingham), whose work provided an 80% completion of the backbone resonance assignments of the SHP2 catalytic domain in the solution state. Mapping of these assignments to the crystal structure of the SHP2 catalytic domain (PDB: 3B7O) [38] revealed a noticeable stretch of unassigned regions correlating with amino acid residues located directly in the phosphatase active site. In particular, those residues residing in the dynamic phosphate binding loop (P-loop), a core component of the conserved protein tyrosine phosphatase (PTP) catalytic motif, exhibited the longest stretch of continuous unassigned resonances. Indeed, several NMR assignment studies on related protein tyrosine phosphatases [205-207] have reported similar

observations of missing assignments owing to the inherent conformational dynamics and structural plasticity of active site loop regions, which serve as critical determinants for governing substrate recognition and accessibility of phosphopeptide ligands to the phosphatase active site. As a crucial structural component of SHP2 ligand recognition, it was deemed prudent to concentrate on attempting to resolve the missing key assignments surrounding the catalytic pocket. Defining a more completely assigned active site would thus enable rapid and accurate prediction of active-site directed inhibitors and fragments binding to the SHP2 catalytic domain in the solution state. As such, the first aim of this chapter was to therefore expand the current assignment coverage by increasing the deuteration levels of the protein sample to improve data quality and thus yield further resolution of amide backbone resonance assignments of unassigned residues lining the SHP2 catalytic pocket.

As a precursor to fragment based library screening to enable the discovery of novel small molecule inhibitors targeting SHP2, the second aim of this chapter was to characterise the structural integrity and solution behaviour of SHP2 catalytic domain in phosphate and HEPES buffer conditions and verify perturbations of these phenomena in the presence and absence of the SHP2 inhibitor, NSC-87877. In obtaining such information, our intention was to characterise the ligand binding behaviour of SHP2 catalytic domain with a putative inhibitor to validate $^{15}N$-$^{1}H$ HSQC NMR spectroscopy as a viable method for mapping and interrogation of inhibitor and fragment binding modes in solution. This approach provides complementary insights into the optimal buffer conditions for small- molecule screening to ensure targeting of a stable, correctly folded and physiologically-relevant protein conformation.

As a final objective, to further explore the ligand binding capabilities of SHP2 in solution, the backbone amide assignments of the SHP2 catalytic domain derived from triple resonance NMR experiments were utilised to map the binding mode of inhibitor NSC87877 and uncover structural and dynamic aspects of inhibitor interactions with the SHP2 catalytic domain that would be directly applicable to the objective of developing novel SHP2 inhibitors. More specifically, binding of the NSC87877 inhibitor may serve to stabilise the active site loop motions in a more rigid conformation allowing additional assignments in these regions of protein structure to be determined, in particular those of the amino acid residues lining the catalytic pocket. This would provide the first experimental validation of SHP2 inhibitor binding by NMR spectroscopy and illuminate key amino acid residues that are directly or indirectly perturbed upon interaction with NSC-87877 and the SHP2 catalytic domain.

## 3.2 Results

In order to guide the development of novel SHP2 inhibitors, the first aim of this chapter of work was to expand the total assignment coverage of the SHP2 catalytic domain by increasing the levels of protein deuteration in an attempt to resolve the previously unassignable residues lining the phosphatase active site. Unfortunately, the restrictions of $^{15}N$-$^{1}H$ protein-observed NMR spectroscopy to the study of proteins with molecular weights <40 kDa prohibited collection of backbone assignments for the 72 kDa full length SHP2 protein. To circumvent this problem, concerted efforts by past and current members of Professor Michael Overduin's laboratory have culminated in the assignment of the individual domains of SHP2 which collectively comprise all of the structured domains in the full length protein.

These efforts have led to the sequential backbone resonance assignments of the n-SH2, c-SH2, tan-SH2 (n-SH2 and c-SH2 domain together) and catalytic domain of SHP2. In particular, 80% of the backbone amide resonances of the SHP2 catalytic domain have been assigned (from the unpublished work of Dr Timothy Knowles, University of Birmingham). These contributions have elevated our understanding of the structure and dynamics of SHP2 in solution and offer a robust platform for interrogating ligand binding to SHP2 using NMR spectroscopy. This platform allows the characterisation of inhibitor interactions with individual domains, and can be exploited for the development of SHP2 inhibitors with novel modes of inhibition.

### 3.2.1 Expression and purification of $^{2}$H-$^{15}$N-$^{13}$C-labelled SHP2 catalytic domain

Prior to conducting triple resonance experiments in an attempt to further resolve unassigned regions of SHP2 catalytic domain, a uniformly-labelled sample of $^{2}$H-$^{15}$N-$^{13}$C SHP2 catalytic domain was grown in 99% $D_2O$ using M9 minimal media supplemented with deuterated glucose to increase deuterium-labelling of the protein to ~85% (see method described in 2.2.5). These conditions were selected and optimised based on previous conditions which used M9 media supplemented with $^{13}$C-labelled glucose to generate a triple-labelled protein sample yielding 80% of backbone amide assignments of the SHP2 catalytic domain (data collected by Dr Timothy Knowles). After obtaining purified protein to ~95% homogeneity, as judged by SDS PAGE analysis (Figures 3.1 & 3.2), $^{2}$H-$^{15}$N-$^{13}$C-labelled SHP2 catalytic domain was incubated with TEV protease to facilitate removal of the His-tag. Analysis of the SDS gel (Figure 3.1B) suggests that cleaved protein elutes predominantly in the wash fraction with negligible levels of detectable protein in the flow-through fraction. Despite the SDS gel indicating a clear mass

difference (an unquantified observation that appears consistent with the removal of a 2.6 KDa $His_{x6}$ tag and linker region) between uncleaved and cleaved protein (Figure 3.1B, lanes 2 and 3, respectively), the virtually undetectable levels of cleaved protein in the flow-through fraction were initially surprising. However, the total volume of pooled protein collected from the initial purification step (which formed the overnight TEV-treated dialysate) was ~ 6 ml suggesting that the majority of the cleaved protein species may have still been loaded onto the 5 ml HisTrap column which had only passed through one column volume. A subsequent wash step with 25 ml wash buffer (5 column volumes) proved sufficient to recover the residual cleaved protein that had yet to elute (Figure 3.1B, lane 5). Thereafter, the recovered cleaved protein was concentrated and dialysed into a final NMR buffer containing 50 mM sodium phosphate, 150 mM NaCl, 0.5 mM TCEP. The final protein concentration was 0.52 mM which was subsequently diluted to 0.5 mM for collection of NMR backbone assignments.

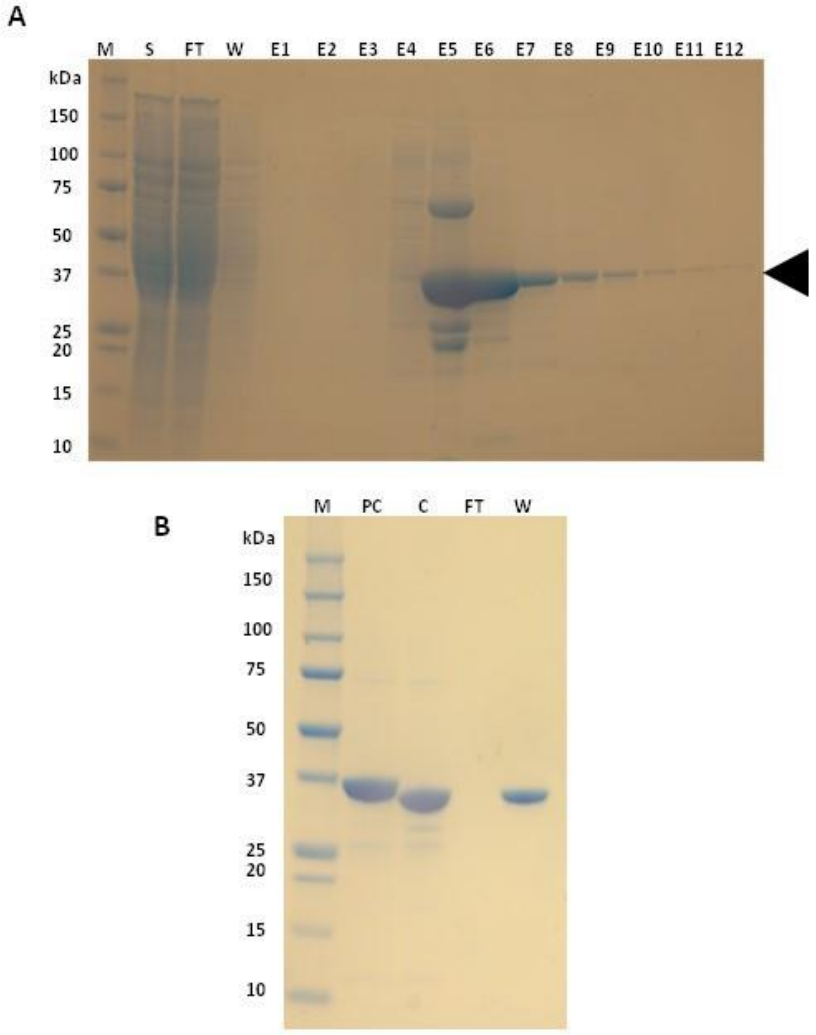



**Figure 3.1 Affinity purification and SDS-PAGE analysis of $^{2}H$-$^{15}N$-$^{13}C$ – labelled SHP2 catalytic domain.** A. From left to right: molecular weight markers (M), supernatant (S), flow-through from soluble lysate fraction (FT), wash (W), elution fractions (E1-E12). Arrowhead indicates overexpression of SHP2 catalytic domain. B. Pooled 1.5 ml elution fractions obtained from affinity purification using a 5 ml HisTrap affinity column (E5-E12) pre and post-incubation with TEV protease (2 mg/ml). From left to right: pooled elution fractions pre-TEV incubation (PC), pooled elution fractions post-TEV incubation (C), flow-through collected from TEV-treated overnight dialysate (FT), wash (W). Analysis of the molecular weight differences between pre and post-TEV incubation fractions indicate the successful removal of $His_{x6}$-tag + linker region. The wash fraction containing TEV-cleaved protein was further purified to homogeneity by size-exclusion chromatography.

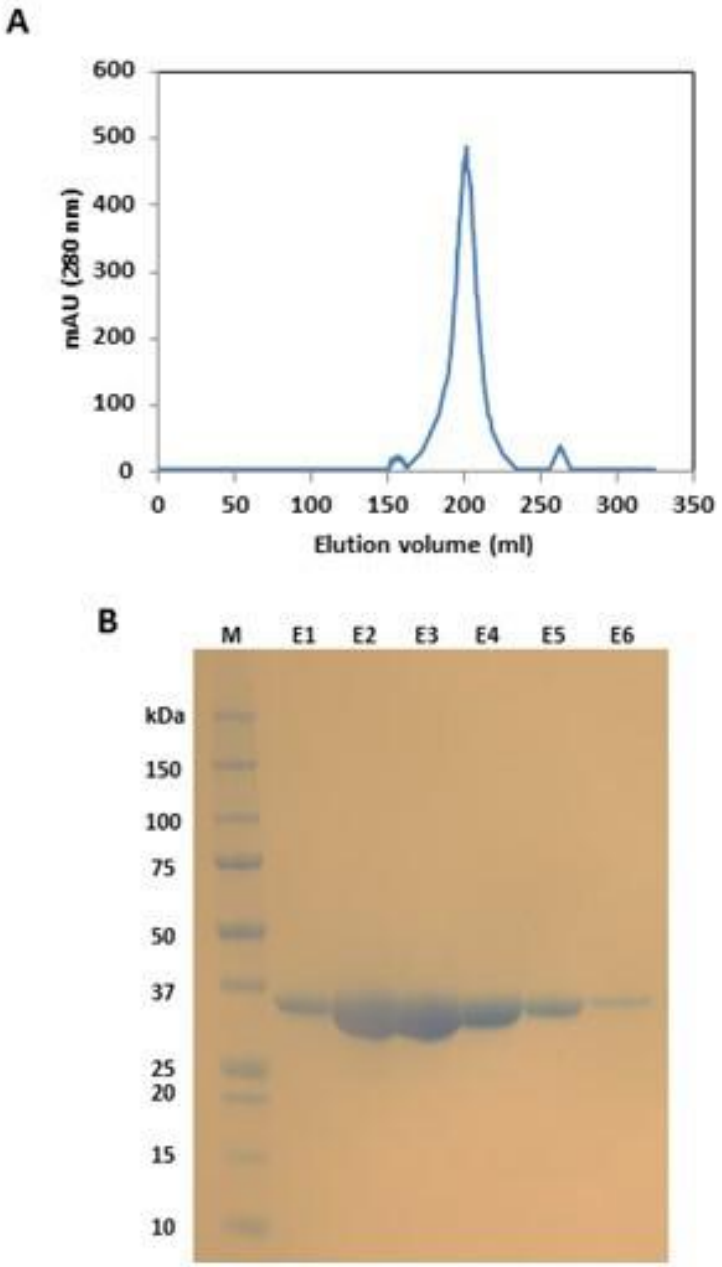



**Figure 3.2 Purification of $^{2}H$-$^{15}N$-$^{13}C$-labelled SHP2 catalytic domain by size exclusion chromatography**. A. Analysis of purified $^{2}H$-$^{15}N$-$^{13}C$–labelled SHP2 catalytic domain by SDS-PAGE and size-exclusion chromatography. TEV-cleaved protein in the wash fraction was concentrated and further purified by size-exclusion chromatography. A. Size exclusion chromatogram showing elution of $^{2}H$-$^{15}N$-$^{13}C$–labelled SHP2 catalytic domain as a single dominant peak indicating the presence of a monomeric species in solution. B. From left to right: molecular weight markers (M), elution fractions from size-exclusion chromatography run (E1-E6).

### 3.2.2 $^{15}$N-$^{1}$H TROSY HSQC of SHP2 catalytic domain in the presence of HEPES

To determine the solution conditions that would yield optimal resolution of backbone amide resonances, and to provide a direct comparison of spectra quality under different solution conditions, we recorded $^{15}$N-$^{1}$H TROSY HSQC spectra of $^{15}$N-labelled SHP2 catalytic domain in the presence of both HEPES and phosphate buffer conditions at the same protein concentration of 0.5 mM. In conditions containing 50 mM HEPES, 150 mM NaCl and 0.5 mM TCEP, only approximately 10% of the expected amide signals in the $^{15}$N-$^{1}$H TROSY HSQC spectra were observed (Figure 3.3A).

### 3.2.3 $^{15}$N-$^{1}$H TROSY HSQC of SHP2 catalytic domain in the presence of phosphate

In contrast to HSQC spectra recorded in HEPES, $^{15}$N-$^{1}$H TROSY HSQC spectra of the SHP2 catalytic domain recorded in the presence of 50 mM phosphate generated spectra of excellent quality with the "fingerprint" of the protein giving good dispersion of backbone amide signals (Figure 3.3B). Importantly, there were no detectable signs of peak clustering in the central region of the spectrum or any other characteristic hallmarks of protein denaturation or unfolding under the experimental conditions tested indicating that SHP2 catalytic domain existed as an appropriately folded species in phosphate buffer conditions.

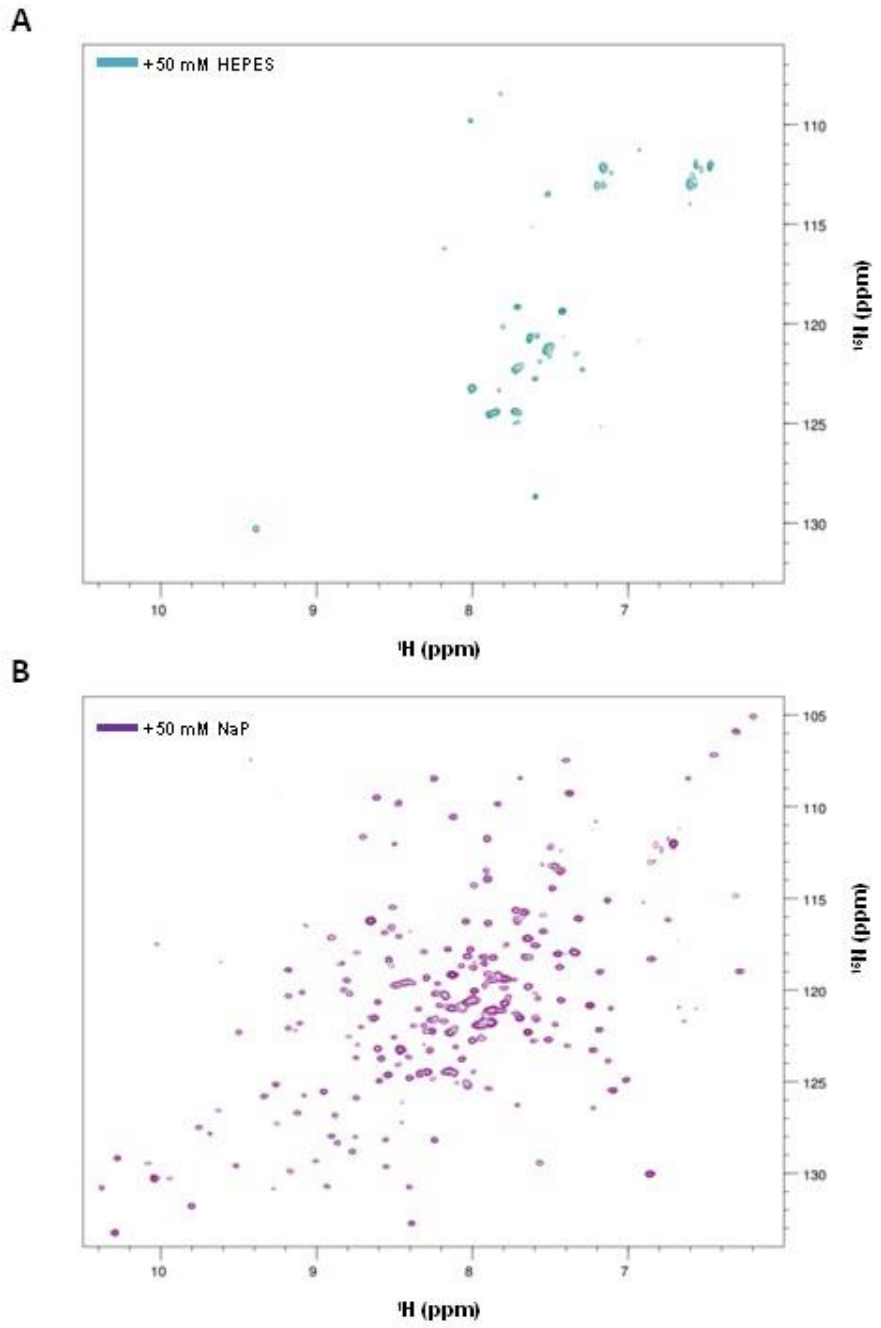



**Figure 3.3. $^{15}$N-$^{1}$H TROSY HSQC spectra of $^{15}$N-labelled SHP2 catalytic domain recorded in the presence of HEPES and phosphate buffer conditions**. A. $^{15}$N-$^{1}$H TROSY HSQC spectra of $^{15}$N-labelled SHP2 catalytic domain recorded in 50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP (turquoise). The recorded spectra yielded a significant reduction in the expected number of backbone amide signals. The marked lack of spectral resolution suggests that under HEPES buffer conditions, SHP2 catalytic domain exhibits prominent structural and conformational plasticity. B. Comparative $^{15}$N-$^{1}$H NMR HSQC spectra of $^{15}$N-labelled SHP2 catalytic domain recorded in 50 mM sodium phosphate buffer (purple). The recorded spectra yielded an even dispersion of backbone amide signals with vast improvement in peak resolution. This indicates that the presence of phosphate imparts structural and conformational stabilisation of SHP2 catalytic domain under these conditions.

### 3.2.4 Refinement of sequential backbone assignments of the SHP2 catalytic domain

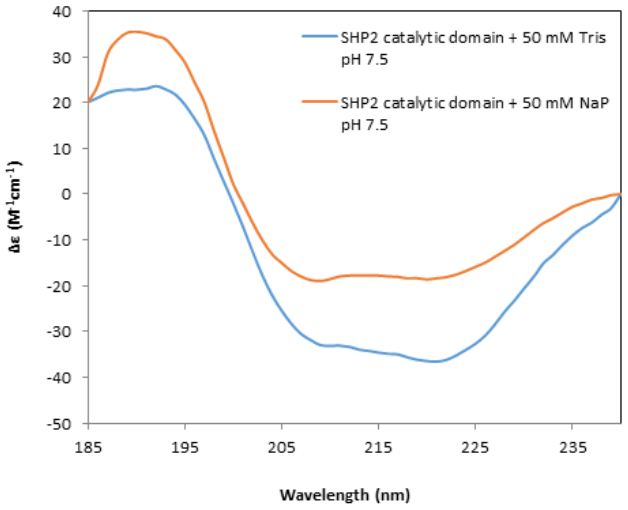
Expression of $^{2}H$-$^{15}N$-$^{13}C$-labelled SHP2 catalytic domain in M9 minimal media supplemented with deuterated glucose was conducted in an attempt to increase deuterium-labelling which serves to reduce the relaxation properties of $^{13}C$ and $^{15}N$ nuclei (through suppressing spin diffusion) and optimise data collection by improving signal to noise. TROSY versions of HNCA, HNCOCA, HNCACB, HNCOCAB, HNCO and HNCACO experiments were conducted in order to facilitate sequential amide backbone assignments on uniformly $^{2}H$-$^{15}N$-$^{13}C$-labelled protein [203]. The backbone assignments were conducted in the presence of 50 mM NaP, 150 mM NaCl and 0.5 mM TCEP, which were identical to the conditions used for collection of the previous assignment dataset (unpublished data collected from Dr Tim Knowles) and known to yield a well dispersed fingerprint of the SHP2 catalytic domain with excellent spectral resolution of backbone amide signals. Picking of amide peaks was initially conducted on $^{15}N$-$^{1}H$ TROSY HSQC spectra with the sequential backbone assignments completed using the CCPN-NMR Analysis software package (see Appendix figure 4). Once completed, the assigned amide peaks in the new dataset were overlaid onto the previous dataset in order to validate the previous set of assignments. From the previous dataset (unpublished data collected by Dr Tim Knowles), approximately 80% of amide backbone assignments (corresponding to ~212 assigned amide peaks from a predicted total of 274). Analysis of this dataset (recorded on ~70% deuterated protein) with the newly collected assignment data recorded on ~90% deuterated protein enabled positive validation of 198 of the previously determined backbone amide assignments, along with the emergence of approximately 10 additional amide peaks which arose presumably as a result of increased deuterium-labelling and thus improved relaxation. Unfortunately, of the 10 additional amide peaks obtained none of these

resonances were assignable perhaps due to low peak intensity. The remaining 20% of unassigned amide resonances are likely due to internal dynamics in the protein leading to broadened linewidths. Unsurprisingly, many of these missing assignments correlate with residues located in exposed loop regions of protein structure. Despite the unsuccessful attempts to obtain additional assignments, enhancing the level of deuterated protein enabled refinement of the previous assignment dataset by altering 14 of these assignments with the correct and less ambiguous resonances.

### 3.2.5 SHP2 catalytic domain is folded in the presence of phosphate

To further interrogate the effect of phosphate on the solution conformation of SHP2 catalytic domain and to provide orthogonal confirmation of a correctly and relevantly folded protein conformation under these conditions, circular dichroism (CD) analysis of SHP2 catalytic domain was conducted in the presence and absence of Tris and phosphate buffers to probe any differences in secondary structure integrity in both conditions. Interestingly, despite significant differences in chemical shift dispersion in $^{15}$N-$^{1}$H TROSY HSQC spectra in the presence of HEPES and phosphate buffer (Figures 3.3A and 3.3B, respectively), circular dichroism analysis revealed that SHP2 catalytic domain exists in a folded conformation in the presence of 50 mM Tris pH 7.5 and 50 mM sodium phosphate pH 7.5. Furthermore, negative CD traces were observed at 222 nm and 208 nm and positive traces observed at 193 nm confirming a predominantly α-helical protein fold (Figure 3.4). Although the geometry of individual α-helices and β-sheets are structurally unperturbed in both conditions (as evidenced by close similarities in the ratios of the different secondary structure elements calculated in DICHROWEB), the difference in CD traces in both the presence and absence of phosphate suggest that phosphate is likely

acting by imparting stabilisation of the overall protein fold in solution. This observation is supported by the marked disparities in backbone amide signal dispersion in $^{15}N$-$^{1}H$ TROSY HSQC spectra for SHP2 catalytic domain in the presence and absence of phosphate (Figure 3.3A and 3.3B), and suggests that the phosphate-dependent stabilisation of SHP2 folding may be representative of a more physiologically relevant conformational state of the protein. Furthermore, the calculated values quantifying the ratio of individual secondary structure elements of the SHP2 catalytic domain in both Tris and HEPES buffer conditions are highly consistent with those reported for the crystal structure of the SHP2 catalytic domain (PDB: 4B7O) [38]. Indeed, the crystal structure of the SHP2 catalytic domain reports a 31% helical content in addition to a 19% β-sheet content, both of which are highly corroborative with the values reported in Figure 3.4 calculated from DICHROWEB secondary structure analysis.




| | α-helix (%) | β strand (%) | Turns (%) | Disordered (%) |
|---|---|---|---|---|
| SHP2 catalytic domain + 50 mM Tris pH 7.5 | 32 | 16 | 18 | 34 |
| SHP2 catalytic domain + 50 mM NaP pH 7.5 | 31 | 16 | 20 | 33 |

**Figure 3.4. Circular dichroism (CD) analysis of unlabelled SHP2 catalytic domain in the presence of phosphate and non-phosphate buffers**. Comparative CD spectra of unlabelled SHP2 catalytic domain (0.1 mg/ml) were recorded both in the presence of Tris buffer pH 7.5 (blue) and sodium phosphate buffer pH 7.5 (red). Negative CD traces recorded for SHP2 catalytic domain at 222 nm and 208 nm, and positive traces recorded at 193 nm suggest a predominantly α-helical protein secondary structure. The contrasting spectral profiles recorded for each buffer condition despite testing at the same protein concentration indicate a difference in the overall protein fold, which provides further evidence of differential stabilisation of the protein in each buffer condition. Buffer subtracted CD traces for SHP2 catalytic domain in the presence of both Tris and phosphate buffer conditions were subsequently analysed using the DICHROWEB server in order to quantify the ratio of secondary structure elements.

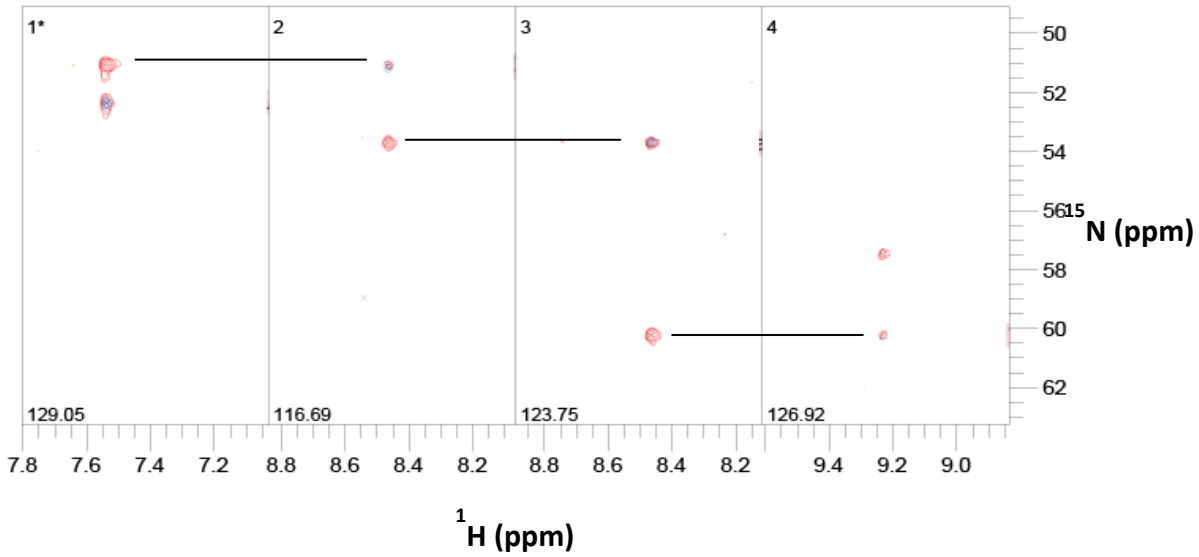



**Figure 3.5. Methodology of sequential NMR backbone resonance assignments of the SHP2 catalytic domain.** Strip plots of a HNCA experiment showing overlapping of i and i-1 residues.

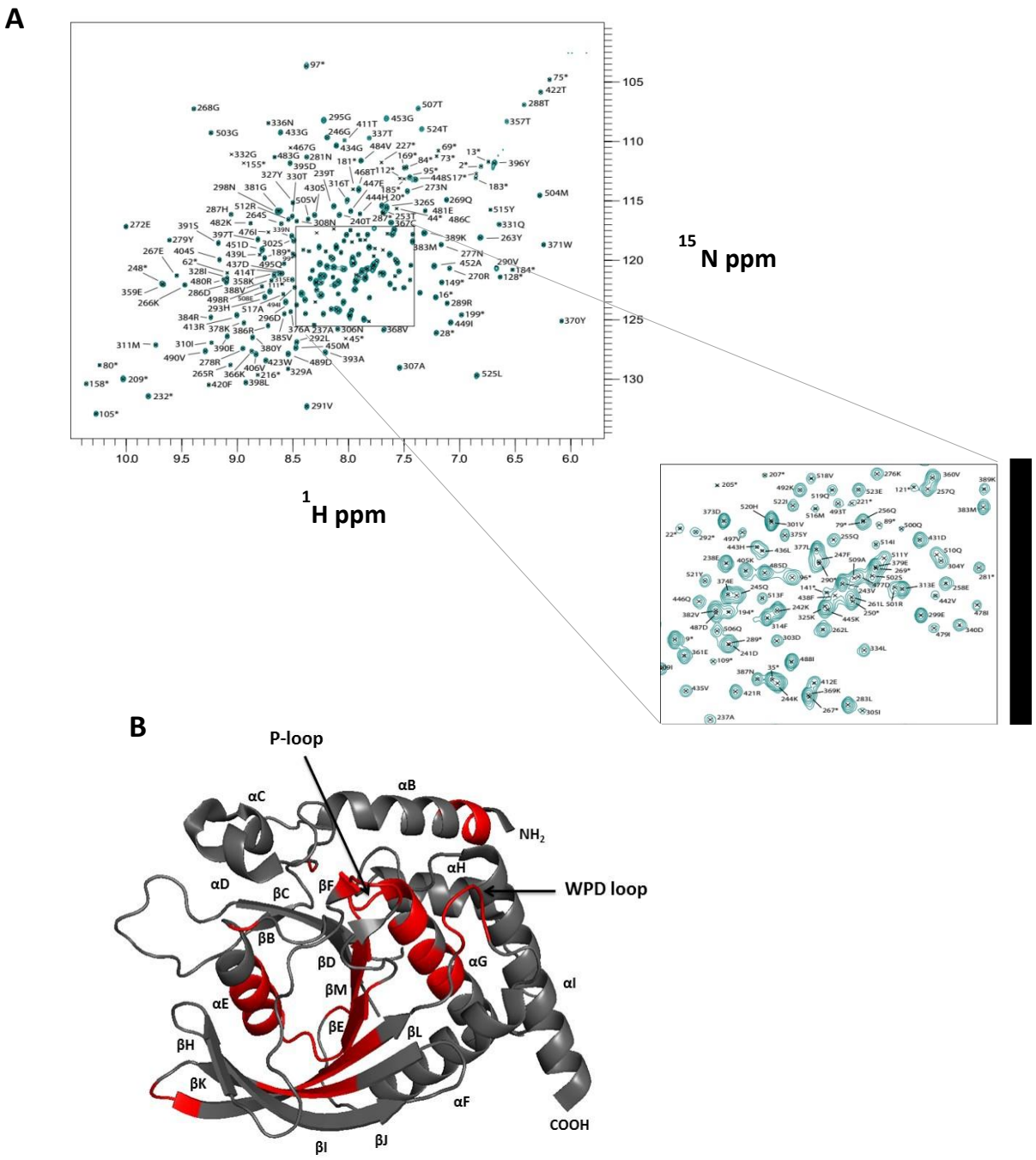



**Figure 3.6. Refinement of NMR backbone resonance assignments for SHP2 catalytic domain by enhanced deuteration**. A. $^{15}N$-$^{1}H$ TROSY HSQC of $^{2}H$-$^{15}N$-$^{13}C$-labelled SHP2 catalytic domain grown in 99% $D_2O$ and recorded in 50 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.5 mM TCEP. Previous backbone assignments determined by Dr Tim Knowles were re-analysed, confirmed and refined using a secondary assignment dataset reported in this study. B. Mapping of the refined NMR backbone resonance assignments to the crystal structure of SHP2 catalytic domain (PDB code: 3B7O). Regions in grey correspond to assigned regions of protein structure. Regions in red correspond to regions of protein structure that remain unassigned. Arrows indicate the assignment state for the conserved PTP catalytic motifs, including the phosphate binding loop (P-loop) and WPD loop.

### 3.2.6 SHP2 catalytic domain exists as a monomeric species in the presence of HEPES

After confirming that the SHP2 catalytic domain existed in a folded conformation in the presence and absence of phosphate, the next aim was to eliminate the possibility that differences in the backbone amide signal dispersion in $^{15}N$-$^{1}H$ TROSY HSQC spectra were a result of aggregation of SHP2 catalytic domain in solution. Since the primary objective of this chapter of work was to utilise $^{15}N$-$^{1}H$ NMR spectroscopy to unravel the structural mechanisms and dynamic insights into SHP2 inhibitor binding as well as characterise the most appropriate conditions permitting interrogation of inhibitor binding, the next aim was to ensure that the phosphate-dependant enhancement of signal resolution in $^{15}N$-$^{1}H$ TROSY HSQC spectra was being recorded on a single monomeric species and that the presence of phosphate was not inducing protein aggregation or multimerisation events that may act to perturb ligand binding. Furthermore, an additional aim was to confirm that the poor $^{15}N$-$^{1}H$ TROSY HSQC spectra recorded in HEPES is due to conformational exchange within the protein and not a result of protein aggregation.

In order to investigate whether the presence of phosphate was alleviating possible aggregation of SHP2 observed in HEPES buffer conditions and elucidate the predominant oligomerisation state of SHP2 in solution, analytical ultracentrifugation (AUC) was conducted on SHP2 catalytic domain in both the presence of HEPES and phosphate buffer. Testing SHP2 catalytic domain at two separate concentrations ($A_{280}$ = ~0.5 and 0.9 AU) in both phosphate and HEPES buffer conditions revealed the presence of a stable and dominant monomeric species in solution with an apparent molecular mass of 34 KDa which is consistent with the expected molecular mass of 34 kDa predicted using ProtParam (Figure 3.7). This observation was also independently confirmed by size-exclusion

chromatography during purification of SHP2 catalytic domain in both phosphate and HEPES buffer where examination of the elution profiles confirmed that SHP2 catalytic domain eluted with a single monomeric peak and existed as a monomeric species in both the presence and absence of phosphate (Figure 3.8). Although calibration of the size-exclusion column with a range of proteins with known molecular weight was not conducted, SDS gels of SHP2 catalytic domain (Figure 3.2B) along with AUC experiments reveal an apparent molecular weight of 34.0 KDa which is identical to the expected molecular weight of the monomer of 34.3 KDa predicted from the protein sequence using the ProtParam program (see Appendix figure 4).

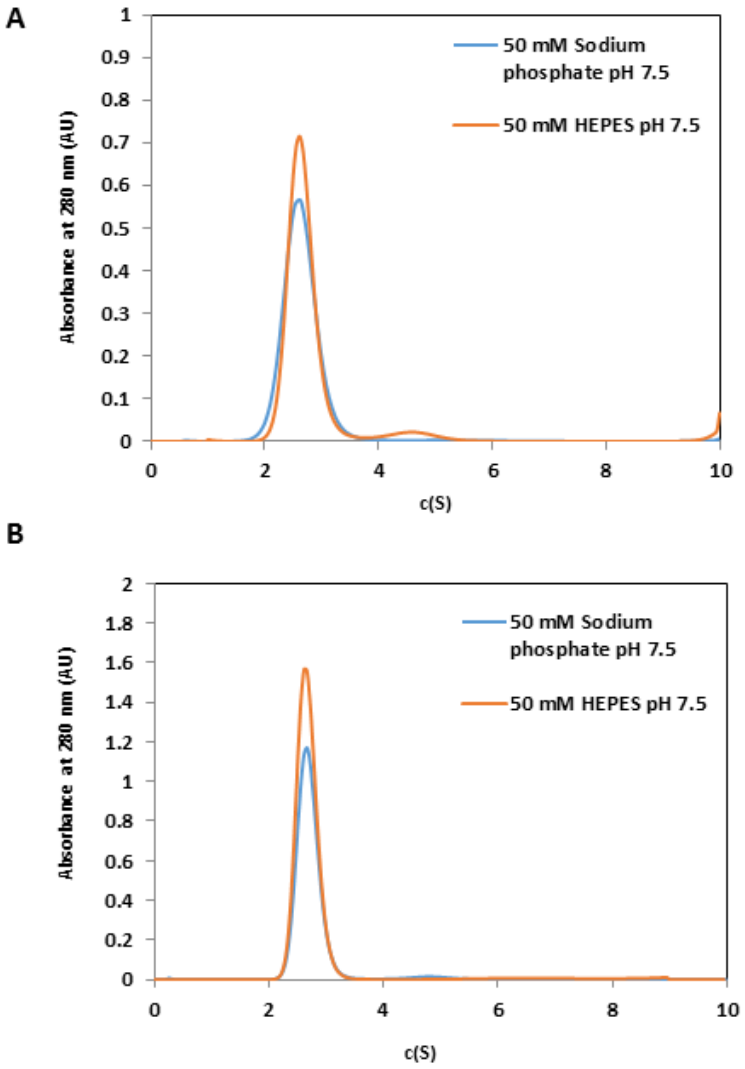



**Figure 3.7. Analytical ultracentrifugation of SHP2 catalytic domain in HEPES and sodium phosphate buffer conditions**. A) AUC of SHP2 catalytic domain ($Abs_{280}$ = ~0.5 AU) in 50 mM HEPES pH 7.5 and 50 mM sodium phosphate 7.5 reveals a homogenous population of monomeric species with an apparent molecular mass of 34.0 KDa which is very close to the expected molecular mass of 34.3 KDa predicted from the protein sequence. B) AUC of SHP2 catalytic domain ($Abs_{280}$ = ~0.9 AU in 50 mM HEPES pH 7.5 and 50 mM phosphate pH 7.5 similarly reveals a homogenous population of monomeric species with an identical molecular mass. No higher order aggregates or oligomeric species were observed.

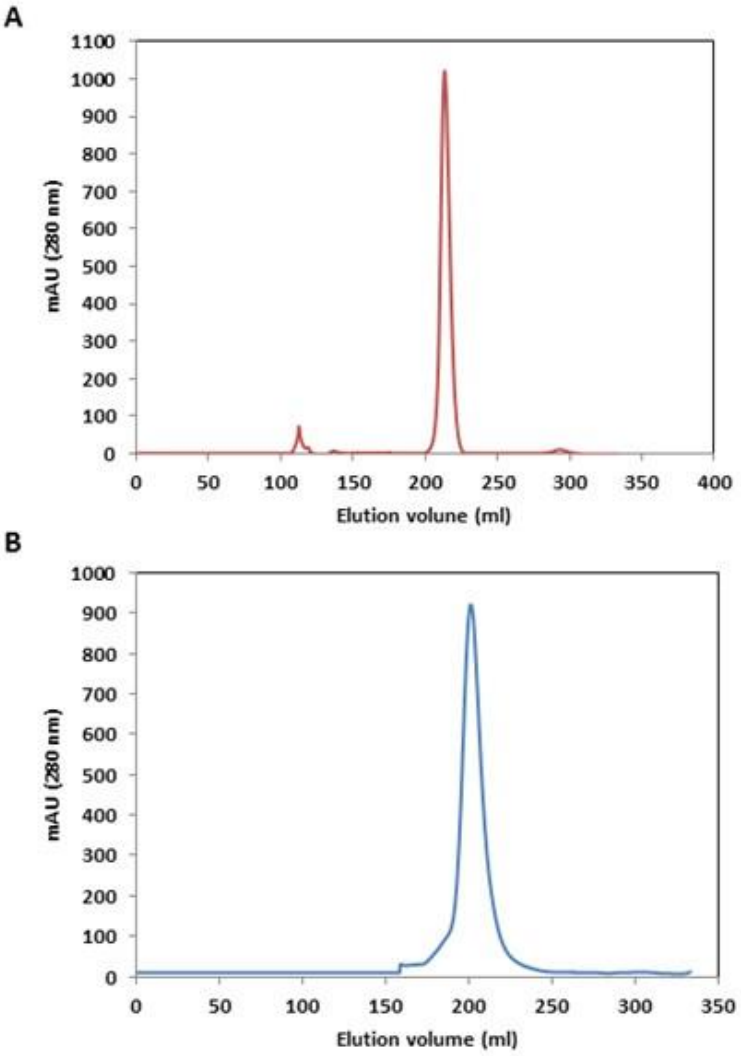



**Figure 3.8. Analysis of SHP2 catalytic domain by size-exclusion chromatography in both HEPES and phosphate buffer conditions**. A. Purification of SHP2 catalytic domain by size-exclusion chromatography in the presence of buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP. B. Purification of SHP2 catalytic domain by size-exclusion chromatography in the presence of buffer containing 50 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.5 mM TCEP. SHP2 catalytic domain elutes as a single dominant peak in both HEPES (red) and phosphate (blue) buffer indicating the presence of a single monomeric species.

### 3.2.7 Sequential titration of $^{15}$N-labelled SHP2 catalytic domain with phosphate

After confirming that SHP2 catalytic domain was appropriately folded and present as a stable monomeric species in the presence of phosphate and HEPES buffer, the next aim was to verify whether the folded conformation of SHP2 catalytic domain was structurally perturbed in the presence of a SHP2-specific inhibitor and whether it was possible to map chemical shift perturbations induced through inhibitor binding under these experimental conditions. Preliminary experiments conducted by titrating promiscuous phosphate inhibitors sodium orthovanadate (an irreversible phosphatase inhibitor) [208] and glycerol-2-phosphate with $^{15}$N-labelled SHP2 catalytic domain and recording $^{15}$N-$^{1}$H TROSY HSQC spectra revealed no obvious indications of inhibitor binding to SHP2 catalytic domain in the presence of phosphate (data not shown), despite sodium orthovanadate being a well characterised phosphatase inhibitor with nanomolar binding affinity. These experiments were also repeated with the validated salicylic acid-based SHP2 inhibitor IIBO8 [143], but no chemical shift perturbations or alternative binding phenomena were observed to indicate an interaction between the inhibitor and the SHP2 catalytic domain, despite the addition of a 5-fold molar excess of inhibitor relative to protein. This observable lack of inhibitor binding was therefore attributed to the compromised solubility of IIB08 in solution combined with the high phosphate concentrations necessary for optimal $^{15}$N-$^{1}$H HSQC spectral resolution.

The initial observations pertaining to a lack of inhibitor binding therefore motivated a search for an alternative SHP2 tool compound with improved solubility for which a binding site could be elucidated by mapping chemical shift perturbations in $^{15}$N-$^{1}$H HSQC NMR spectra causing previously collected backbone assignments for the SHP2 catalytic

domain. After a thorough investigation of the literature the commercially available disulphonic acid-based inhibitor NSC-87877 was identified as a promising candidate which exhibited an IC50 of 0.3 μM in a biochemical pNPP assay. Furthermore, a combination of in-silico modelling and mutagenesis studies suggest that NSC-87877 is an active site-directed inhibitor of the SHP2 catalytic domain, but structural evidence confirming the proposed mode of inhibition has remained elusive [136].

Earlier observations regarding the lack of inhibitor binding by $^{15}N$-$^{1}H$ NMR spectroscopy led to the hypothesis that in the presence of 50 mM phosphate concentration the phosphatase active site was saturated resulting in conditions that actively disfavour rather than promote the binding of active site-targeting inhibitors and leading to an improve backbone amide signal resolution in $^{15}N$-$^{1}H$ HSQC spectra. This is consistent with previous observations where attempts to observe inhibitor binding were unsuccessful for sodium orthovanadate and compound IIB08, both of which are validated inhibitors targeting the phosphatase active site.

After reasoning that the high phosphate concentration could be acting to occlude the SHP2 active site and therefore perturb inhibitor binding, as a precursor to conducting $^{15}N$-$^{1}H$ HSQC titrations against SHP2 catalytic domain with compound NSC-87877, it was suggested that employing a sequential phosphate titration against $^{15}N$-labelled SHP2 protein in 50 mM HEPES pH 7.5 was a necessary step in order to determine the optimum phosphate concentration that yielded sufficient $^{15}N$-$^{1}H$ HSQC spectral resolution but also maximised the possibility of observing inhibitor binding by outcompeting the active site-bound phosphate. Increasing equivalents of sodium phosphate buffer (pH 7.5) to final concentrations of 10 mM and 20 mM were then added and then a series of $^{15}N$-$^{1}H$

SOFAST-HMQC experiments of $^{15}$N-labelled SHP2 catalytic domain were recorded in the absence of compound NSC-87877 in order to gauge whether backbone amide signals were sufficiently resolved and that conditions were conducive for inhibitor NSC-87877 titrations (Figure 3.11).

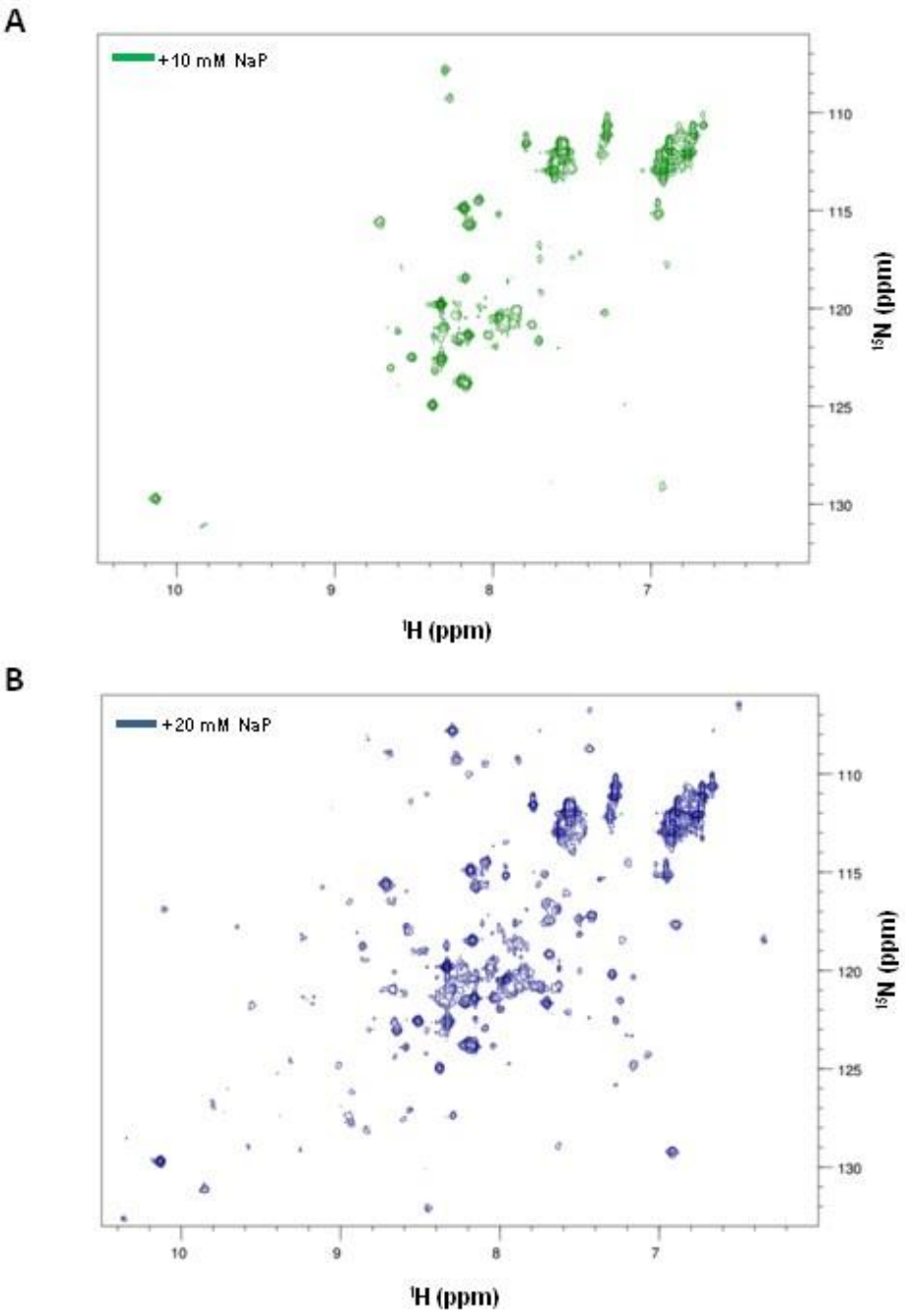



**Figure 3.9. $^{15}$N-$^{1}$H SOFAST-HMQC NMR spectra of $^{15}$N-labelled SHP2 catalytic domain in the presence of 10 mM and 20 mM sodium phosphate pH 7.5**. A) Titration of $^{15}$N-labelled SHP2 catalytic domain in 50 mM HEPES buffer pH 7.5 with 10 mM sodium phosphate buffer pH 7.5. Despite showing a clear improvement in spectral resolution, $^{15}$N-$^{1}$H SO-FAST HMQC spectra recorded in 10 mM phosphate show insufficient resolution of backbone amide resonance signals for inhibitor titrations. B) Secondary sequential titration of 15N-labelled SHP2 catalytic domain with an additional equivalent of 10 mM sodium phosphate buffer pH 7.5 yielding a final concentration of 20 mM sodium phosphate. Titration of an additional equivalent of sodium phosphate pH 7.5 to a final concentration of 20 mM yields marked improvement in spectral resolution that was deemed conducive for inhibitor titrations.

Although $^{15}N$-$^{1}H$ SOFAST-HMQC spectra recorded in 10 mM phosphate provided minor improvements in backbone amide signal resolution (Figure 3.9A), the observed improvement in spectral resolution was deemed insufficient for chemical shift perturbation analysis. However, increasing the phosphate concentration to 20 mM yielded sufficient improvement in spectral resolution to enable chemical shift mapping with peak dispersion comparable to the $^{15}N$-$^{1}H$ TROSY HSQC spectra recorded in the presence of 50 mM phosphate (Figure 3.3B). Despite observing improvements in spectral resolution at a phosphate concentration of 20 mM, $^{15}N$-$^{1}H$ SOFAST-HMQC spectra yielded noticeable broad linewidths owing to the relatively large molecular weight of the SHP2 catalytic domain, which typically occupies the upper molecular weight limit for the study of protein structures by protein-observed NMR methods. To address the caveat of broad linewidths and achieve overall enhancement of spectral resolution and sensitivity, $^{15}N$-$^{1}H$ TROSY spectra for $^{15}N$-labelled SHP2 catalytic domain were recorded at a phosphate concentration of 20 mM. The resulting $^{15}N$-$^{1}H$ TROSY HSQC experiment yielded spectra of excellent resolution with sufficiently dispersed backbone amide peaks (Figure 3.10). The marked improvement in spectral quality and clarity enabled the employment of titrations of $^{15}N$-labelled SHP2 catalytic domain with inhibitor NSC-87877 to probe inhibitor binding by monitoring perturbations of backbone resonances in the presence and absence of inhibitor.

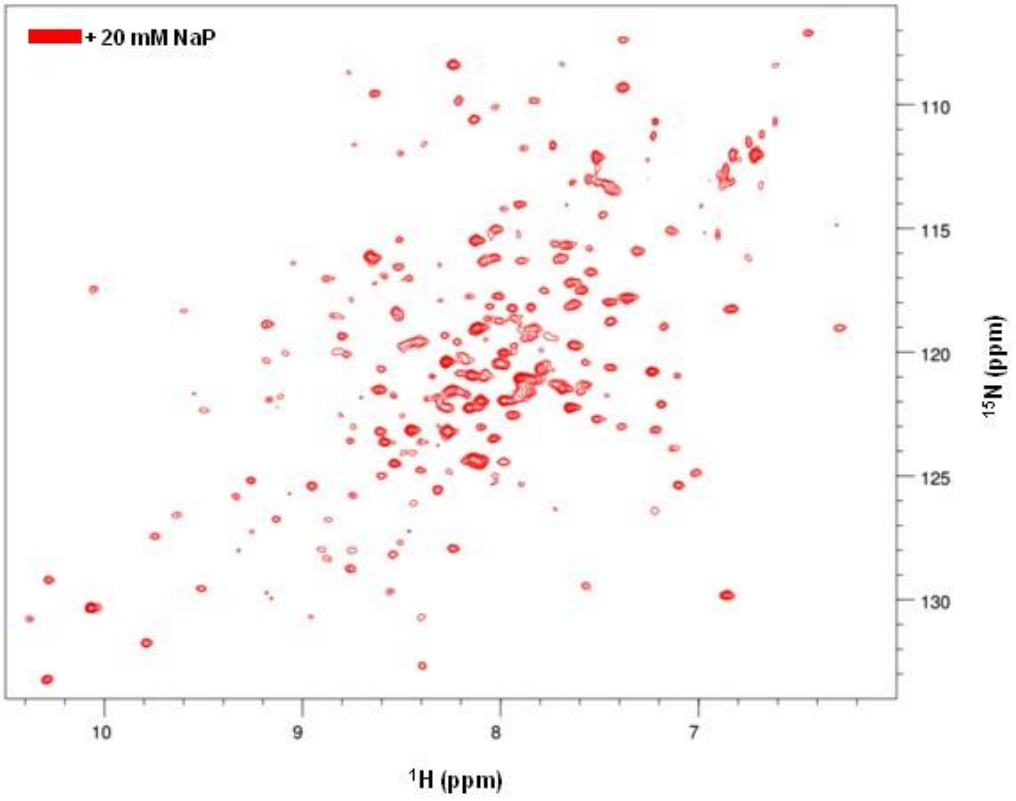



**Figure 3.10. $^{15}$N-$^{1}$H TROSY HSQC NMR spectra of $^{15}$N-labelled SHP2 catalytic domain in the presence of 20 mM sodium phosphate pH 7.5.** $^{15}$N-$^{1}$H TROSY HSQC experiments on $^{15}$N-labelled SHP2 yielded reduced line broadening and marked improvement in overall spectral resolution and sensitivity. This enabled conditions that were conducive for NSC-87877 inhibitor titrations.

### 3.2.8 Titration of $^{15}$N-labelled SHP2 catalytic domain with compound NSC87877

Addition of a 2-fold equivalent of NSC-87877 (600 µM) relative to the concentration of $^{15}$N-labelled SHP2 catalytic domain (300 µM) in the presence of 20 mM sodium phosphate pH 7.5 revealed no obvious indications of compound precipitation or insolubility and the compound appeared highly soluble in 5% (vol/vol) DMSO compared to the previously tested IIB08 inhibitor. Overlaying $^{15}$N-$^{1}$H TROSY HSQC spectra of $^{15}$N-labelled SHP2 catalytic domain recorded in 20 mM sodium phosphate pH 7.5 in the presence and absence of inhibitor NSC-87877 revealed no observable chemical shift perturbations of backbone amide signals. However, closer inspection of $^{15}$N-$^{1}$H TROSY HSQC spectra revealed an unexpected disappearance of particular amide resonances upon addition of the NSC-87877 inhibitor, which initially suggested an intermediate exchange mechanism of inhibitor binding. At this point, an alternative but equally plausible explanation for this observation was also considered, suggesting that addition of the NSC-87877 inhibitor reintroduces some of the intermediate exchange dynamics into the SHP2 catalytic domain through liberation of bound phosphate. After confirming the disappearance of a total of five peaks upon addition of NSC-87877, the identity of these peaks was verified by mapping to the corresponding resonance assignments in an attempt to generate the first structural evidence of the binding mode of NSC-87877 to the SHP2 catalytic domain (Figure 3.11A) Interestingly, all five of the disappearing peaks exhibited assignable backbone resonances which corresponded to residues Y279, N281, T357, T468 and T507. Mapping these residues to the crystal structure of the SHP2 catalytic domain (PDB: 3B7O) revealed that all of these residues are located in and around the active site cavity of SHP2 catalytic domain, confirming a possible active site-directed mechanism of inhibition (Figure 3.11B)

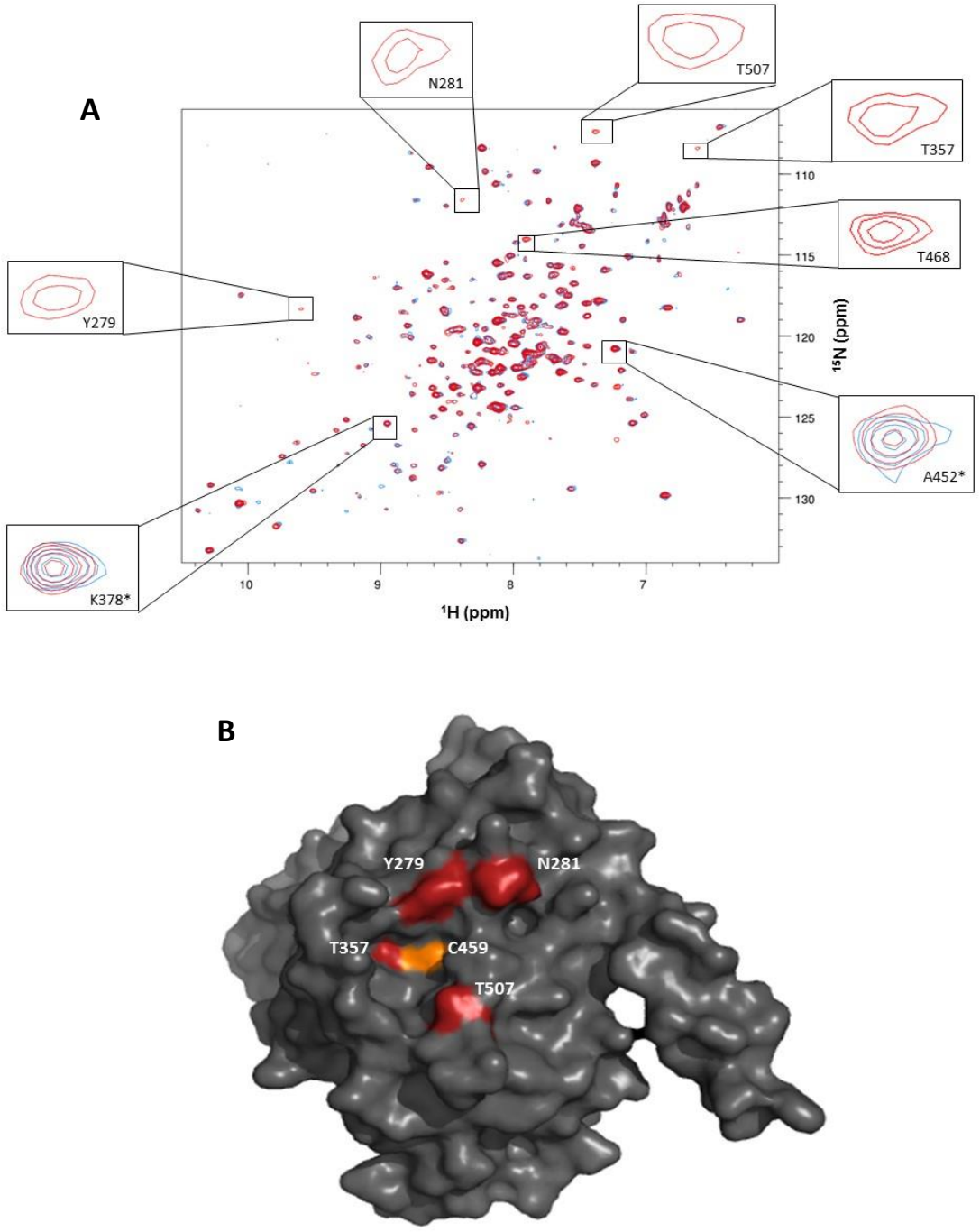



**Figure 3.11 Titration of $^{15}$N-labelled SHP2 catalytic domain with inhibitor NSC-87877**. A. Overlay of $^{15}$N-$^{1}$H TROSY HSQC spectra of 300 µM $^{15}$N-labelled SHP2 catalytic domain in the presence (light blue) and absence (red) of 600 µM NSC-87877. Backbone resonances of assigned residues Y279, N281, T357, T468 and T507 for disappearing peaks have been magnified for clarity, along with control peaks K378* and A452*. B. Surface representation of SHP2 catalytic domain highlighting disappearing peaks upon addition of NSC-87877 (red) and the catalytic cysteine 459 (orange). Perturbed residue T468 is not shown as residue is buried in the interior of the protein and proved difficult to project using a surface representation of the SHP2 catalytic domain.

## 3.3 Discussion

In order to assist the discovery and development of more efficacious SHP2 inhibitors and characterise the ligandable potential of the SHP2 catalytic domain under solution conditions, triple-resonance NMR methods were employed in conjunction with $^{15}N$-$^{1}H$ TROSY HSQC experiments in an attempt to assign previously unresolvable backbone resonances in the active site loop regions of SHP2 catalytic domain, enabling exploitation of structural and dynamic mechanisms of SHP2 inhibitor binding as a precursor to fragment-based ligand discovery.

NMR spectroscopy provides an indispensable tool for the characterisation of molecular structure and offers powerful structural and dynamic insights into protein-ligand recognition events in the solution state. The acquisition of structural information by NMR-based methods typically begins with elucidation of resonance assignments for each backbone amide group in a particular protein target which can then be utilised for rapid determination of inhibitor binding to active and regulatory sites on the protein. This allows for molecular interactions and conformational flexibilities of macromolecular structures to be characterised under physiological conditions and at levels of resolution sufficient for confirming small-molecule binding to a given target and guiding inhibitor development as part of NMR-based ligand discovery campaigns.

Prior to commencing this work, the elucidation of a previous set resonance assignments revealed an approximate 80% coverage of backbone amide signals for the SHP2 catalytic domain (unpublished data generated and analysed by Dr Tim Knowles and Dr Mark Jeeves). Interestingly, further interrogation of the assignment data identified approximately 20% of the observed resonances were still unassignable despite increasing deuterium-

labelling and, after mapping to the crystal structure of SHP2 (PDB: 3B7O), these were found to be predominantly situated in the dynamic loop regions comprising the phosphatase active site. These regions that exhibit a lack of assignable resonances can be rationalised by considering multiple explanations, which include the intrinsic flexibility and structural plasticity of the phosphatase active site loop regions leading to rapid interconversion between multiple conformational ensembles, and chemical exchange mechanisms operating in the phosphatase active site leading to perturbations in signal intensity.

In order to circumvent these obstacles and attempt to resolve the hitherto uncharacterised backbone resonances for residues lining the active site loops of SHP2, the level of protein deuteration was increased by supplementing M9 minimal growth media with deuterated glucose. This was in contrast to the methodology used for sample preparation for collection of the previous assignment dataset (recorded by Dr Tim Knowles) which used $^{13}$C-labelled glucose. For the SHP2 catalytic domain which comprises a molecular weight of 36 kDa, it was suggested that enhancement of protein deuteration could yield a favourable strategy for resolving backbone resonances associated with the dynamic and catalytic components of SHP2 structure. Typically, in order to overcome the difficulties associated with NMR data collection on larger protein structures (i.e. for proteins exceeding a molecular weight >20 kDa), deuterium-labelling is commonly employed to improve signal-to-noise by suppressing spin diffusion and reducing the overall relaxation properties of $^{13}$C and $^{15}$N-labelled nuclei [209]. Unfortunately, despite continued attempts, efforts to obtain any additional assignments of the SHP2 catalytic domain were unsuccessful, with the majority of the active site loop assignments remaining largely unassigned. However, the new triple resonance dataset recorded on >90% deuterium-

labelled protein was successfully utilised in order to refine the initial round of assignments. This enabled the correction and reassignment of around 20 previously unclear and ambiguous backbone resonances.

Despite unsuccessful attempts in gaining additional assignments in the phosphatase active site by enhancing protein perdeuteration, difficulties associated with assigning flexible regions of protein structure is not an uncommon occurrence and has been reported by previous phosphatase NMR assignment studies, particularly those conducted on PRL-3. For example, determination of the NMR-solution structure of PRL-3 revealed that despite possessing overall similar folds to PTEN and PAC-1, the loop regions comprising the respective active sites all exhibit significant structural plasticity [205]. The variation in conformational flexibility of phosphatase active site loops has been suggested to arise from the difference in side chain distances between residues Cys 104, Arg 110 (both in the PTP catalytic signature motif HCXXGXXR) and Asp 72 in PRL-3 which in comparison to the corresponding residues in PTEN are separated by a greater distance. This has been suggested to arise largely from the key position of the Asp residue which serves as the general acid in the conserved catalytic mechanism of all phosphatases. The extent of conformational flexibility exhibited by active site loop regions across the phosphatase superfamily has been similarly confirmed by high throughput crystallography studies reporting significant conformational propensities of loop regions from a comprehensive analysis of 25 phosphatase crystal structures [38].

A necessary prerequisite before initiating NMR-based ligand discovery is ensuring the macromolecular target of interest has undergone robust characterisation in solution in order to determine optimum conditions for preserving protein folding, stability and

catalytic integrity. This also includes a thorough evaluation of structural and dynamic perturbations of the target in the presence of putative ligands or inhibitors. Collectively, these preliminary measures ensure that validated hits emerging from NMR-based small molecule screening assays are binding to a physiologically relevant and stable form of the protein, rather than preferentially binding to an unfolded or oligomerised state of the target. In this way, as a precursor to small molecule screening a primary objective was to elucidate the optimum screening conditions under which SHP2 catalytic domain would exhibit the relevant folding and stability representative of physiological conditions. Indeed, attempts to characterise the preferred solution conditions for optimal stability of the SHP2 catalytic domain raised interesting questions concerning the role of phosphate in stabilising the structure of SHP2 catalytic domain, as evidenced by the marked differences in amide signal resolution in protein-observed NMR experiments conducted in the presence and absence of phosphate. In parallel with observations from AUC confirming that SHP2 catalytic domain existed as a stable monomeric species in the presence and absence of phosphate, $^{15}$N-$^{1}$H HSQC spectra independently confirmed that the presence of phosphate induced a more conformationally stable protein structure. For the purposes of NMR assignment determination and the use of assigned backbone resonances to characterise ligand binding sites, these experiments could only be conducted in the presence of phosphate due to the greater resolution and dispersion of the amide signals in comparison to the profoundly unresolved spectra recorded in the absence of phosphate. Indeed, our initial attempts to map chemical shift perturbations by $^{15}$N-$^{1}$H TROSY HSQC experiments in the presence and absence of SHP2 inhibitors sodium orthovanadate and IIB08 revealed a surprising lack of inhibitor binding. In the case of compound IIB08, overlaying $^{15}$N-$^{1}$H TROSY HSQC spectra in the presence and absence of inhibitor revealed no observable

evidence of ligand binding despite a thorough assessment of backbone resonances for chemical shift perturbations or line broadening. However, a likely explanation underlying the lack of IIB08 inhibitor binding concerns the poor solubility of the inhibitor which may have been exacerbated by the requirement for the high inhibitor concentrations (>300 μM) necessary for protein-observed NMR HSQC experiments. Aside from problems associated with compound solubility, it was strongly suspected the large molar excess of phosphate required for sufficient resolution of backbone amide signals posed additional complications for observing ligand binding. This is especially true given that the reported $K_D$ of 5.5 μM for compound IIB08 [143] compared with a $K_D$ of approximately 10-20 mM for phosphate should easily enable compound IIB08 to outcompete the presence of active site bound phosphate. However, this is not true if the binding of phosphate in the active site is significantly tighter than the binding of phosphate acting to stabilise and rigidify the overall protein conformation. Although concentrations of phosphate exceeding 20 mM contribute favourably to the structural and dynamic stabilisation of SHP2 in solution, such conditions are also likely to disfavour inhibitor binding through occlusion and saturation of the phosphatase active site. Despite the likely probability that abrogation of inhibitor binding is a direct result of phosphate-mediated occlusion of the active site, we strongly suspect that the presence of phosphate may also be acting to stabilise the conformational dynamics of SHP2 through binding to multiple different sites on the protein. This is supported by previous experiments where addition of increasing equivalents of the active site directed inhibitor IIB08 (to a final concentration of 2 mM) to $^{15}$N-labelled SHP2 catalytic domain in the absence of phosphate causes minor improvements in peak recovery (presumably through stabilisation of the active site) but is not sufficient to fully reproduce the effects recorded in phosphate which yields excellent dispersion of amide resonance

signals in $^{15}$N-$^{1}$H TROSY HSQC spectra. This leads us to hypothesise that the presence of phosphate could be driving stabilisation of the SHP2 catalytic domain through binding to multiple different sites on the protein which collectively lead to a more favourable conformational ensemble in solution. This would allow for greater stability, folding and resolution of the signature amide signals needed for determination of backbone assignments.

Our hypothesis asserting that the vast molar excess of phosphate necessary for full resolution of the backbone amide signals acts to occlude the phosphatase active site posed significant obstacles for characterising the ligand binding behaviour of the SHP2 catalytic domain in solution. Indeed, this was a particularly striking observation, especially given that both sodium orthovanadate and IIB08 have both been reported to target the phosphatase active site. Although repeated attempts to map the binding site of inhibitor IIB08 by $^{15}$N-$^{1}$H HSQC NMR were unsuccessful, we extended our assessment of SHP2 ligandability by $^{15}$N-$^{1}$H HSQC methods by partially mapping the previously undetermined binding mode of the more potent SHP2 inhibitor, NSC-87877.

As backbone amide resonances are exquisitely sensitive to perturbations in the local chemical environment, the utility of protein-observed NMR spectroscopy provides a rapid method for examining the interactions in solution between small-molecule compounds and $^{15}$N-labelled SHP2 catalytic domain by monitoring chemical shift perturbations of backbone amide resonances present in $^{15}$N-$^{1}$H TROSY HSQC spectra.

Although compound NSC-87877 displayed comparatively greater solubility compared with inhibitor IIB08, efforts to directly map progressive chemical shift perturbations to observe the exchange dynamics of NSC-87877 binding were largely unsuccessful.

Interestingly however, overlaying of $^{15}$N-labelled SHP2 $^{15}$N-$^{1}$H TROSY HSQC spectra in the presence and absence of a 2-fold excess of NSC-87877 inhibitor (600 µM) relative to protein concentration (300 µM) resulted in noticeable disappearance of several backbone amide resonances. The specific changes occurring in HSQC spectra upon addition of a bound ligand (i.e. chemical shift perturbations or line broadening) are governed by the exchange rates between bound and unbound species relative to the difference in chemical shift between bound and unbound states. A progressive decrease in signal intensity or complete disappearance of a resonance signal implies that the exchange rate between the bound and unbound species is in an intermediate exchange regime with respect to the NMR timescale of the experiment, or that we have reverted to a localised intermediate exchange mechanism within the protein seen in the absence of phosphate. In our case, it would therefore be reasonable to suggest that the interaction between compound NSC-87877 and SHP2 catalytic domain is obeying intermediate exchange dynamics, resulting in complete disappearance of specific resonance signals.

As mentioned previously, chemical shift mapping analysis of disappearing peaks to the assigned backbone resonances uncovered a total of five amino acid residues in close proximity to and in the vicinity of the phosphatase active site. These residues include Y279, N281, T357, T468 and T507 (T468 not shown as due to lack of visibility in the surface representation model of SHP2 catalytic domain). Interestingly, although the interaction of NSC-87877 with SHP2 catalytic domain lacks the robust structural evidence required for unravelling precise modes of inhibitor binding, in-silico docking studies predict that the B-ring sulphonic acid moiety of inhibitor NSC-87877 forms hydrogen bonds with the backbone amide group of Arg 465 [136] (Figure 3.12B) – a highly conserved residue in the PTP signature motif located at the base of the phosphatase catalytic cleft [21].

Furthermore, the A-ring sulphonic acid moiety was predicted to form hydrogen bonds with the side chain $NH_3$ group of K280 and the side chain $NH_2$ group of N281. Both K280 and N281 are non-conserved PTP residues located adjacent to the key phosphotyrosine recognition loop [136]. Indeed, studies investigating the role of residues K280 and N281 in determining NSC-87877 inhibitor sensitivity found that SHP2 K280V and K280R/N281D mutants show an approximate 3-fold reduction in biochemical activity in a pNPP assay, with $IC_{50}$ values of 1.1 and 1.0 µM for SHP2V280 and SHP2 K280R/N281D, respectively, compared to the 0.3 µM $IC_{50}$ determined for wild-type SHP2. Although no chemical shift perturbations or peak disappearance for backbone resonance signals corresponding to K280 were observed, perturbed residues were mapping to Y279 and N281 upon NSC-87877 addition suggesting that residue N281 and either Y279 and/or K280 form key interactions with the A-ring sulphonic acid group of inhibitor NSC-87877. This provides the first structural validation in support of NSC-87877 binding in which N281 is suggested to play a key role in anchoring the disulphonic acid moiety to the phosphatase active site.

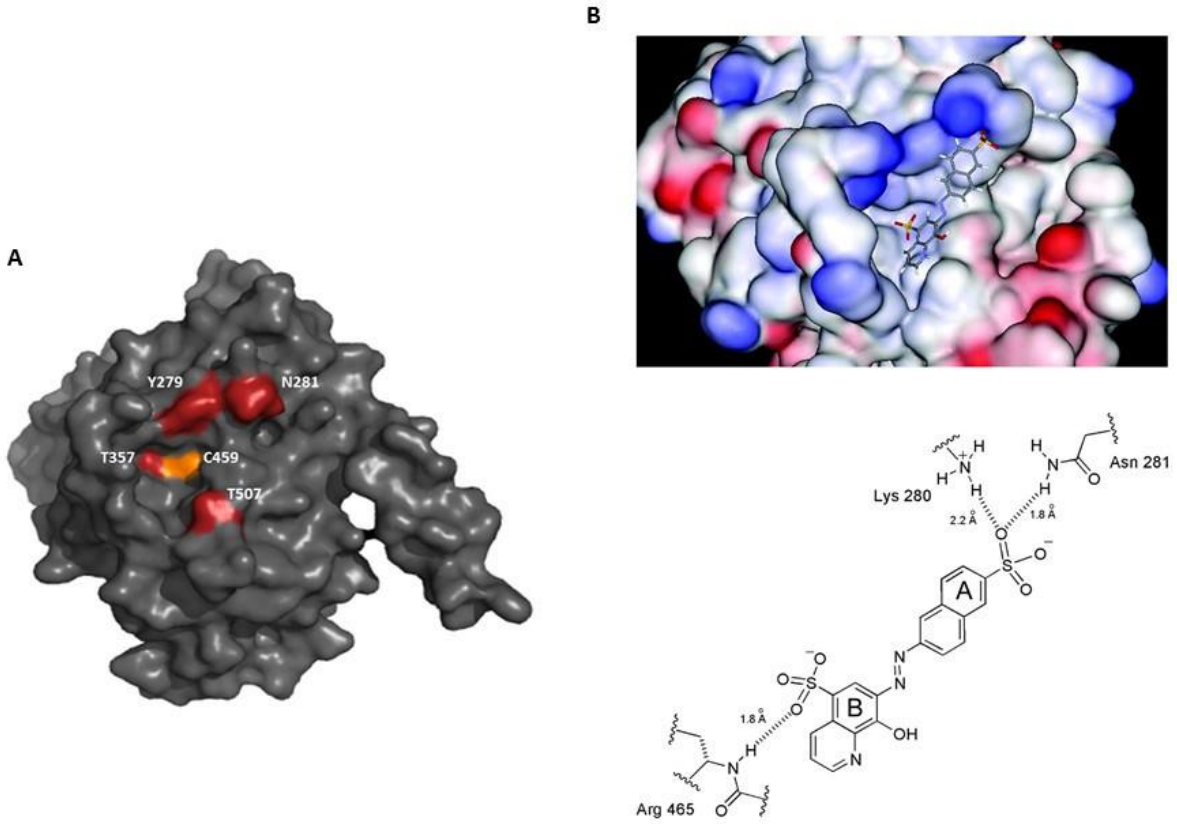



**Figure 3.12 Comparison of inhibitor NSC-87877 binding to the SHP2 catalytic domain by $^{15}$N-$^{1}$H HSQC NMR spectroscopy vs in-silico analysis of the NSC-87877 binding mode.** A. Mapping of assigned residues to the crystal structure of SHP2 that showed perturbation upon addition of NSC-87877 as recorded by $^{15}$N-$^{1}$H HSQC NMR spectroscopy. Interestingly, upon inhibitor NSC-87877 binding, perturbed residues Y279 and N281 reported in $^{15}$N-$^{1}$H TROSY HSQC NMR spectra corroborate with predicted residues K280 and N281, reported from in-silico and mutational studies suggesting that these residues are important for recognition of the A-ring sulphonic acid moiety of NSC-87877. Furthermore, perturbed residues T357 and T507, although not highlighted in the in-silico model of NSC-87877 binding to SHP2 catalytic domain, may form interactions with the sulphonate and hydroxyl moieties either side of the B-ring sulphonic acid moiety. Although not highlighted on the surface representation of SHP2 catalytic domain, due to deep burial in the core of the SHP2 catalytic domain, residues T468 may also play an additional role by forming an interaction with the B-ring sulphonic acid moiety of NSC-87877 in conjunction with Arg 465. B. In-silico analysis of NSC-87877 binding reported by Chen et.al [136] predicting key hydrogen bond interactions between residues K280, N281 of the A-ring sulphonic acid moiety and between R465 of the B-ring sulphonic acid moiety of inhibitor NSC-87877.

The complete disappearance of chemical shift resonances upon addition of NSC-87877 in this manner is therefore likely driven by structural rearrangement of the SHP2 catalytic domain in response to NSC-87877 binding, generating a more conformationally mobile species. These experimental observations have led us to hypothesise that the disappearance of backbone resonance signals arises from the displacement of active site-bound phosphate upon NSC-87877 binding. Furthermore, we propose that binding of NSC-87877 alone is not sufficient to mimic the stabilising contribution of phosphate and thus upon phosphate displacement leads to transient structural flexibility in SHP2 catalytic domain which is manifested by a complete and unrecoverable loss of specific backbone resonance signals. Backbone resonance signals for which we observe a complete loss of peak intensity are therefore likely to correspond to residues occupying dynamically perturbed regions of protein structure that are greatly influenced by the stabilising effect of phosphate and the structural perturbations that occur when phosphate is displaced upon compound NSC-87877 binding. Similar intermediate ligand exchange mechanisms have also been highlighted by other NMR-based studies focussing on inhibitor-binding to HIV-1 protease [210].

## 3.4 Conclusion

In conclusion, the work presented in this chapter offers insights into the molecular structure and conformational dynamics of the SHP2 phosphatase domain under solution conditions, and in the broad context of NMR-based drug discovery provides a robust and informative platform for probing small molecule inhibitor interactions with catalytic sites, or novel allosteric pockets and regulatory sites that may be of key relevance to guiding more potent SHP2 inhibitor development. As such, in the context of this work, compiling a near complete map of the peptide backbone assignments of SHP2 catalytic domain would serve as an excellent prelude for validating hits from fragment-based screening that allows us to interrogate the binding locations of fragments through mapping chemical shift perturbations which serve to illuminate novel allosteric binding sites.

# CHAPTER 4: BIOPHYSICAL AND BIOCHEMICAL VALIDATION OF SHP2 TOOL COMPOUNDS

## 4.1 Experimental aims

After exploring the solution behaviour of SHP2 catalytic domain in the presence of phosphate and non-phosphate buffer conditions, it was clear that phosphate plays a key role in stabilising the overall conformational dynamics of the SHP2 catalytic domain. Biophysical characterisation of SHP2 structure in solution, including AUC and CD analysis combined with size-exclusion chromatography suggested that the protein also exists as a stable, folded and monomeric species in HEPES buffer and in the absence of phosphate, despite yielding poorly resolved $^{15}$N-$^{1}$H HSQC spectra in HEPES.

To further explore the interactions of SHP2 ligands with the SHP2 catalytic domain and provide orthogonal validation of SHP2-ligand binding by alternative methods, a robust panel of biophysical and biochemical assays, was used to complement fragment-based screening. The aim was to characterise the biochemical activities and binding affinities of previously reported SHP2 inhibitors, PHPS1, IIB08 and NSC-87877. Although similar biochemical characterisation of SHP2 inhibitors has previously reported, corroborative biophysical evidence confirming the affinities of these interactions with the SHP2 catalytic domain has remained surprisingly elusive.

To interrogate and validate the binding interactions of inhibitors PHPS1, IIB08 and NSC-87877 with the SHP2 catalytic domain, the biophysical assay panel included ligand-observed NMR experiments (STD and waterLOGSY) and SPR. This was accompanied by an assessment of the effect of these inhibitors on the biochemical activity of SHP2 using a

pNPP activity assay. Together, this approach enabled direct observation of the performance of each inhibitor in each of these assays and permitted selection of the most active inhibitors as validated chemical tools for fragment-based screening. The more robustly validated SHP2 tool compounds emerging from these assays could then be utilised as positive control ligands to authenticate fragment-based screening assays, providing robust confirmation of positive hit matter for further optimisation, which might otherwise prove difficult in the absence of a validated control compound.

## 4.2 Results

In order to validate and profile the activity of known SHP2 inhibitors as a precursor to fragment-based screening, we surveyed the scientific literature and identified a number of suitable candidates for evaluation across a panel of biophysical and biochemical assays, including NMR (STD-NMR and waterLOGSY), SPR and enzyme inhibition assays. Inhibitors selected for validation included compounds PHPS1 [140], IIB08 [143] and NSC-87877 [136] (Figure 4.1), all of which were shortlisted based on their previous disclosure as specific inhibitors of the SHP2 catalytic domain. These tool compounds have all been reported to possess similar binding mechanisms with differential levels of evidence supporting an active-site targeted interaction with the SHP2 catalytic domain. More specifically, compound IIB08 represents the only inhibitor for which an active-site binding mode has been experimentally validated by a co-crystal structure with the SHP2 catalytic domain. While in-silico modelling and enzymatic-inhibition assays have predicted active-site binding modes of interaction for compounds PHPS1 [140] and NSC-87877 [136] the availability of similarly robust structural evidence supporting these mechanisms remain elusive.

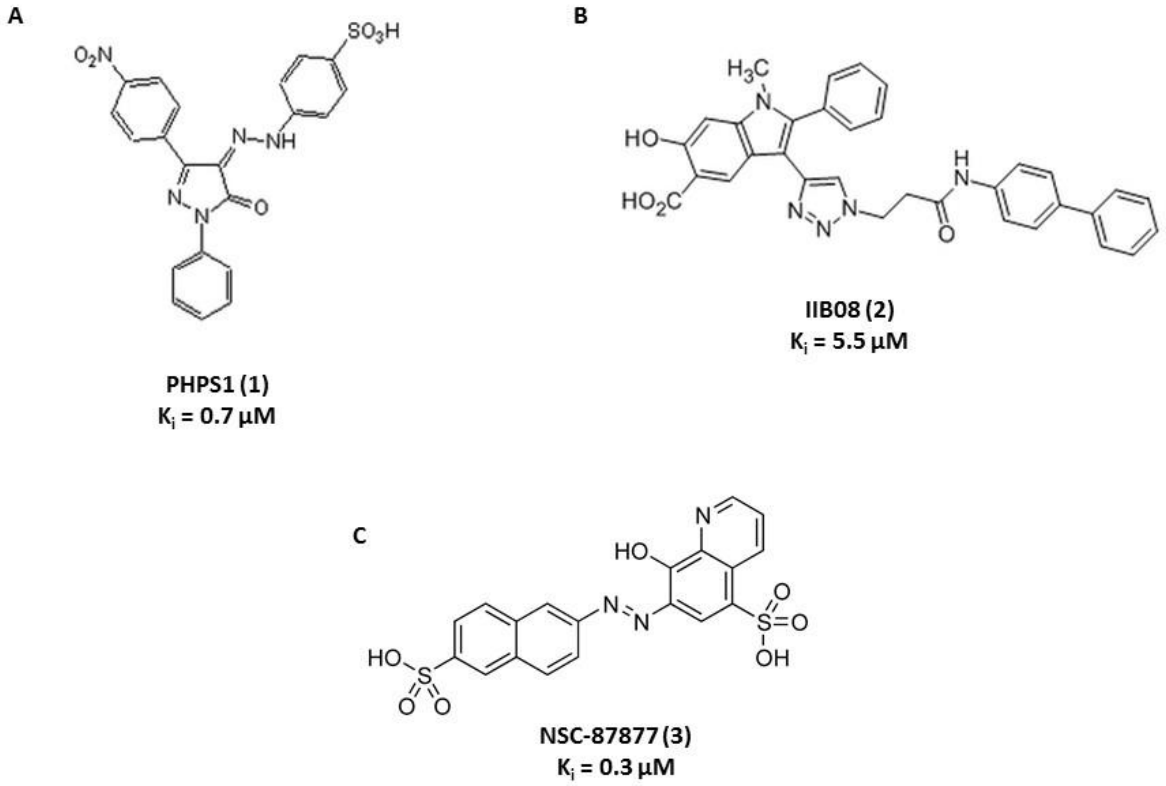



**Figure 4.1 Chemical structures of literature reported SHP2 inhibitors PHPS1, IIB08 and NSC-87877**. A) Chemical structure of compound PHPS1. B) Chemical structure of IIB08. C) Chemical structure of NSC-87877. The inhibitor constants derived from biochemical activity assay measurements are reported as 0.7 μM (PHPS1), 5.5 μM (IIB08) and 0.3 μM.

### 4.2.1 $^{15}$N-$^{1}$H NMR HSQC titrations of SHP2 inhibitors PHPS1 and IIB08

Having previously observed evidence of a putative interaction between compound NSC-87877 and $^{15}$N-labelled SHP2 catalytic domain by $^{15}$N-$^{1}$H HSQC NMR (Chapter 3), it was deemed necessary to perform similar $^{15}$N-$^{1}$H TROSY HSQC titration experiments to see whether it was possible to observe binding of inhibitors PHPS1 and IIB08 by directly probing backbone $^{15}$N-$^{1}$H backbone amide perturbations in recorded $^{15}$N-$^{1}$H TROSY HSQC spectra in the presence and absence of each inhibitor. A thorough survey of the literature revealed that no previous attempts had been made to validate the interaction of compounds PHPS1 and IIB08 with the SHP2 catalytic domain by protein-observed NMR methods. However, unfortunately, despite repeated attempts to improve compound solubility, the relative insolubility of these compounds at the high concentrations required for $^{15}$N-$^{1}$H NMR HSQC experiments in 5% (v/v) DMSO meant that we were unable to map any chemical shift perturbations to confirm the binding of these compounds by protein-observed NMR.

### 4.2.2 SPR characterisation of inhibitor binding to SHP2 catalytic domain

To further validate the interactions of these inhibitors with SHP2 catalytic domain, SPR was utilised to enable direct observation of inhibitor binding in real time, and quantify the binding affinities for the interactions of inhibitors PHPS1, IIB08 and NSC-87877 with the SHP2 catalytic domain. In corroborative fashion to observations made in the $^{15}$N-$^{1}$H HSQC NMR experiments, the lack of solubility of PHPS1 in 5% (v/v) DMSO made it increasingly difficult to measure binding affinities for the interaction with the SHP2 catalytic domain by SPR. However, we were successful in deriving binding affinities for compound NSC-87877 and compound IIB08. Compound NSC-878777 was tested across

an 8-point 5-fold dilution series from 100 µM to the lowest concentration of 0.0064 µM and IIB08 across an 8-point 2-fold dilution series from 250 µM to the lowest concentration of 3.9 µM. In order to derive an appropriate $K_D$ value for compound NSC-87877 and compound IIB08 binding to SHP2 catalytic domain, the resulting binding profiles were fitted to a simple 1:1 Langmuir binding model (using the Biacore evaluation software) [211]. Importantly, fitting the data to this model involves two fundamental assumptions. Firstly, the analyte (compounds NSC-87877 and IIB08) must be homogenous and bind with a 1:1 stoichiometry with the immobilised ligand (i.e. form a monovalent interaction). The same is also true regarding the ligand which is also assumed to be homogenous. The second important consideration of the Langmuir model is the assumption that all binding events are independent.

Derivation of the $K_D$ for the binding interactions between inhibitors NSC-87877 and compound IIB08 binding to SHP2 catalytic domain was then obtained by non-linear curve fitting of the data to the Langmuir equation:

$$R_{eq} = \frac{[L]R_{max}}{[L]+K_D}$$

Where $R_{eq}$ is the response (measured in RU) obtained from binding measured at equilibrium (i.e. the steady state phase of the binding interaction), $R_{max}$ is the maximum response (measured in RU) recorded at saturating concentrations of ligand, [L] is the free ligand concentration, and the $K_D$ is the dissociation constant of the binding interaction measured at equilibrium.

Fitting of the NSC-87877 binding data to the Langmuir model generated a $K_D$ value of 3 μM (with the error in $K_D$ being +/- 0.3 μM) and, to our knowledge, represented the first attempt to characterise the binding of NSC-87877 to SHP2 using SPR (Figure 4.2). Specific binding of compound NSC-87877 to the SHP2 catalytic domain was further confirmed by derivation of an $R_{max}$ (equivalent to the maximum binding capacity of the surface based on the amount of protein immobilised) value which after fitting the data proved to be just underneath the theoretical $R_{max}$ based on the protein immobilisation level.

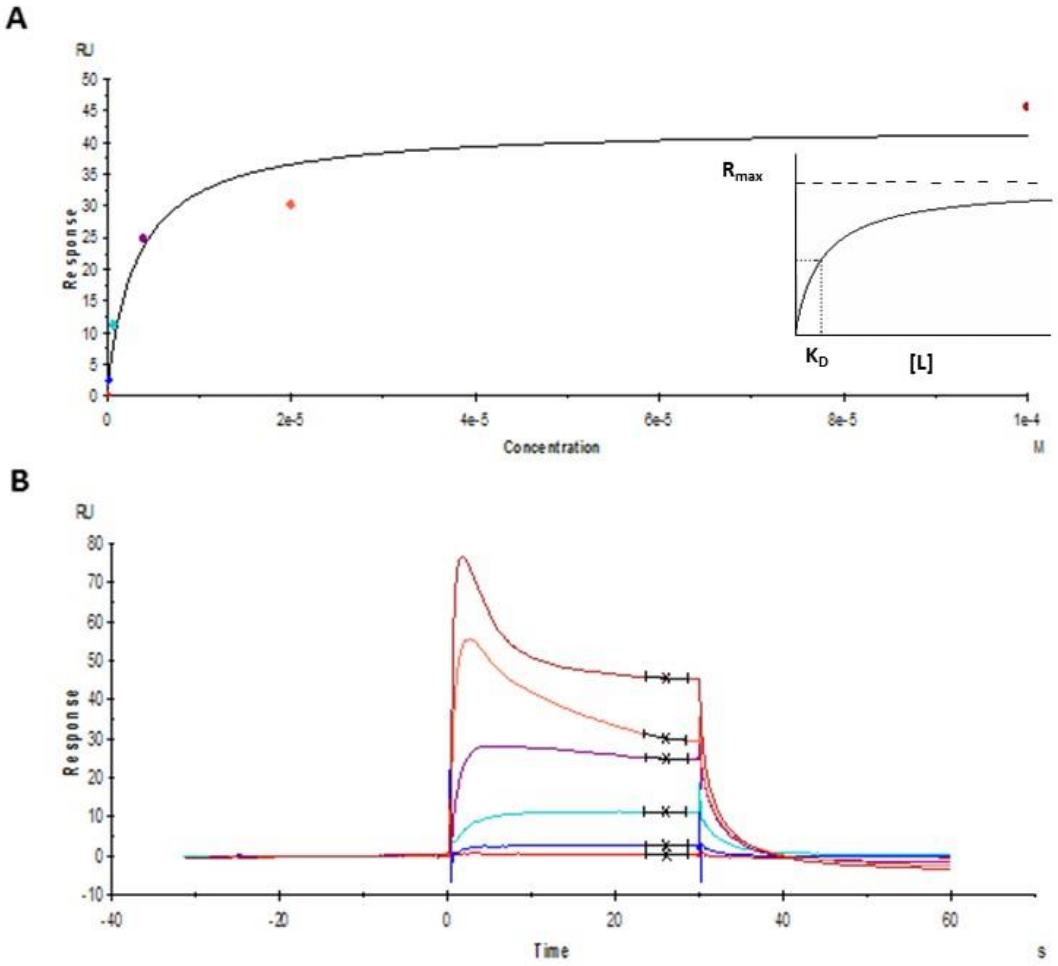



**Figure 4.2 SPR characterisation of compound NSC-87877 binding to SHP2 catalytic domain.** A. An SPR concentration-response curve showing a dose-dependent interaction between compound NSC-87877 and immobilised SHP2 catalytic domain. A 100 μM stock solution of compound NSC-87877 was serially diluted with SPR running buffer across a 5-fold dilution series to give a concentration range of 0.0064 μM – 100 μM. Saturation of the signal response can be observed at 100 μM concentration. The figure inset represents a standard ligand binding curve demonstrating saturation binding behaviour. The $R_{max}$ can be calculated from the asymptote of the curve, and the $K_D$ is defined as the ligand concentration giving half $R_{max.}$ The $K_D$ is measured at equilibrium using the data points at the end of the binding plateau. B. Real-time observation of the increase in signal response upon addition of increasing concentrations of NSC-87877 inhibitor. The return of the response signal to baseline after each inhibitor binding event indicates a specific interaction devoid of aggregation. All response curves were solvent-corrected, and subtracted from the blank (non-immobilised) reference cell and buffer.

Although derivation and quantification of a binding affinity was possible for compound IIB08 ($K_D$ = 3 μM when fitted to a 1:1 Langmuir binding model), closer analysis of the binding profiles suggested that the effects observed were analogous to a compound exhibiting a non-specific binding profile (Figure 4.3). This is supported by the lack of dose-dependent binding in response to increasing concentrations of inhibitor which should lead to the eventual saturation of the signal as all protein sites on the SPR chip gradually become occupied. The relatively unstable response signal which shows a period of signal loss before inhibitor dissociation also indicates that the response signal shown is unlikely to be that of a true protein-ligand interaction. This suggests that the increase in response signal is likely to have arisen as a result of compound aggregation which gives a change in mass and thus refractive index at the surface. The most feasible explanation for this phenomenon is therefore a result of non-specific binding of compound IIB08 which accumulates on the SPR sensor chip surface, yielding a false-positive result.

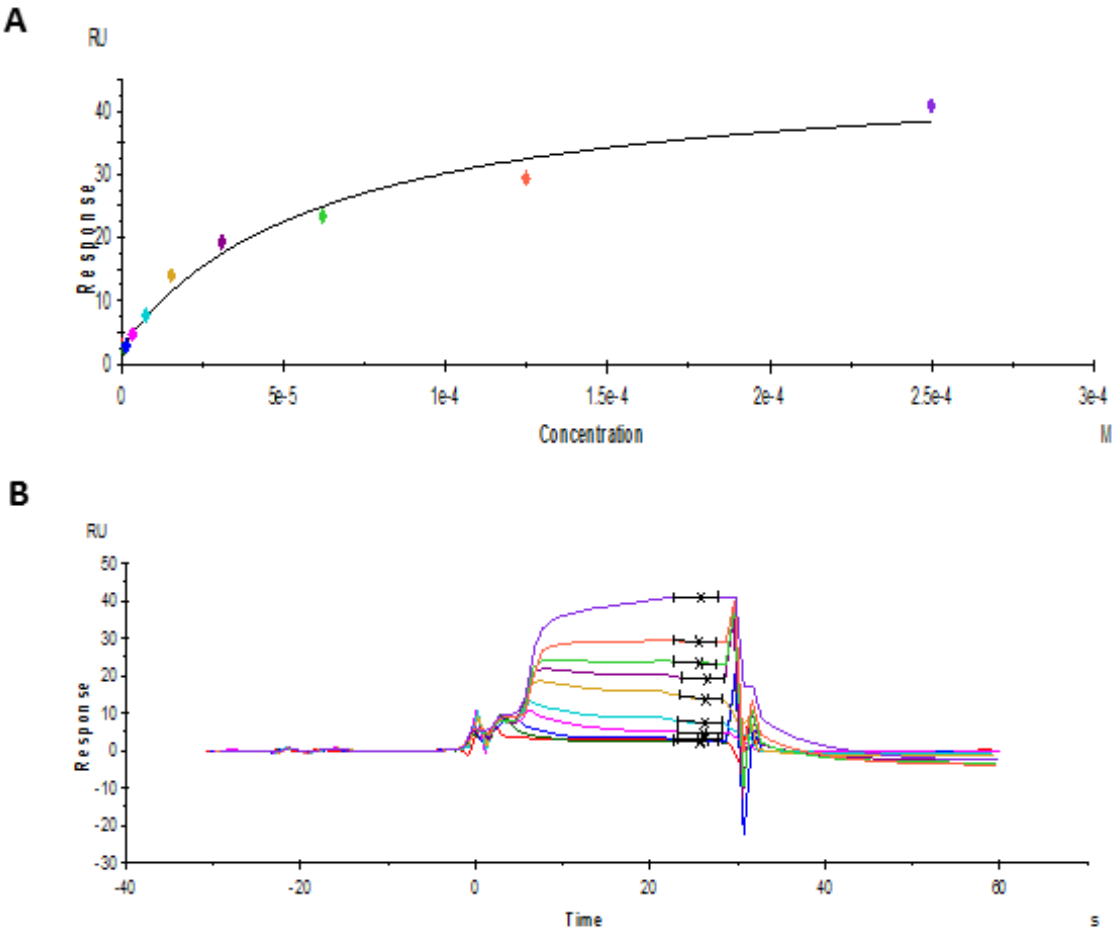



**Figure 4.3 SPR characterisation of non-specific binding of inhibitor IIB08 to the SHP2 catalytic domain.** A. SPR concentration-response curve showing the non-specific interaction between compound IIB08 and immobilised SHP2 catalytic domain. A 100 μM stock solution of compound IIB08 was serially diluted with SPR running buffer across a two-fold dilution series to give a concentration range of 3.9 μM – 250 μM. B) The real-time observation of an increase in signal response upon addition of increasing concentrations of IIB08 inhibitor suggests an aggregation-driving mechanism of compound binding. The slow duration of the response signal in returning to baseline after each binding event provides further evidence of a non-specific interaction. All response curves were solvent-corrected, and subtracted from the blank (non-immobilised) reference cell.

### 4.2.3 Validation of compound NSC-87877 binding by STD-NMR

After confirming binding of compound NSC-87877 and observing non-specific binding of compound IIB08 by SPR, it was deemed necessary to proceed by investigating whether similar observations were reproducible by orthogonal ligand-observed NMR-methods, including STD-NMR and waterLOGSY. Prior to running ligand-observed NMR experiments, 1H-NMR spectra was collected for compounds IIB08 and NSC-87877 to assess compound integrity and probe solubility. $^{1}$H-NMR spectra recorded for compound NSC-87877 enabled successful assignment of the 10 predicted aromatic proton signals (the exchangeable protons associated with the two sulphonic acid groups and the sole hydroxyl group are assumed to be exchanging very rapidly with the bulk water and were not observed). Assignment of the individual proton environments present in compound IIB08 proved extremely challenging due to poor spectra quality which was attributed to a lack of compound solubility. In order to confirm this observation, further interrogation by $^{1}$H-NMR, using a compound with high solubility at known concentrations as an internal control, revealed that the solubility limit of IIB08 was around 50 µM in 5% DMSO (data not shown). Interestingly, this also supported initial observations of suspected compound insolubility in the SPR assays, which generated bi-phasic binding profiles comprising a phase of initial specific binding, followed by progressive non-specific binding as the compound concentration exceeded 50 µM.

To finalise the biophysical validation of compounds IIB08 and NSC-87877, the interactions between inhibitors NSC-87877 and IIB08 with the SHP2 catalytic domain were assessed using ligand-observed methods STD-NMR and waterLOGSY. Interestingly, no observable enhancement of the STD signal or inversion of the waterLOGSY signal was

apparent to confirm an interaction between compound IIB08 and the SHP2 catalytic domain. This was unsurprising as accurate assignment of protein-bound ligand signals is only possible from well resolved $^{1}$H-NMR spectra for the specific ligand of interest. For compound NSC-87877, a very weak STD signal was observed confirming a binding interaction with the SHP2 catalytic domain (Figure 4.4). Surprisingly, despite the observation of a positive hit for compound NSC-87877 by STD-NMR, we were unable to observe a signal inversion in the waterLOGSY spectra to confirm NSC-87877 binding to SHP2 catalytic domain, despite numerous attempts at optimising mixing time parameters to aid waterLOGSY signal resolution. The weak STD spectra and absence of observable waterLOGSY signals were reasoned to be a likely consequence of the preferential sensitivity of these techniques for detection of weak (often mM to high μM) affinity interactions. This is particularly true in light of the $K_D$ value determined for compound NSC-87877, which at 3 μM affinity would have a low off-rate relative to the time-frame of the STD-experiment, suggesting a possible explanation underlying the weakly observed STD signal for NSC-87877 binding to the SHP2 catalytic domain.

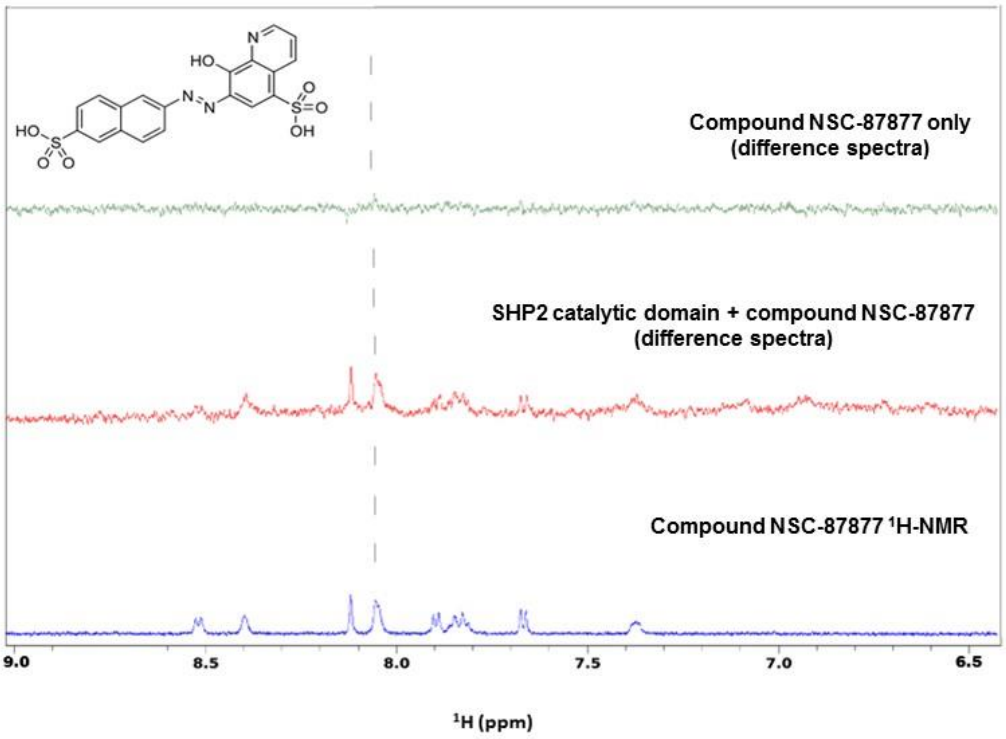



**Figure 4.4 Interaction of inhibitor NSC-87877 with the SHP2 catalytic domain by STD-NMR**. Initial $^{1}$H-NMR spectra was recorded on compound NSC-87877 (100 μM in 5% DMSO) to confirm compound identify and account for all proton environments in the molecule (blue). STD experiments to elucidate an interaction between the SHP2 catalytic domain and compound NSC-87877 were conducted by recording off-resonance (at 25 ppm, without protein saturation) and on-resonance (via selective saturation of the protein at -0.5 ppm) spectra for SHP2 catalytic domain (5 μM) in the presence of compound NSC-87877 (100 μM). The red spectra represents the difference spectra obtained by subtracting the on-resonance spectra from the off-resonance spectra enabling identification of the NSC-87877 compound signals receiving saturation transfer from binding to the SHP2 catalytic domain. The green spectra represent the difference spectra recorded for compound NSC-877877 only (in the absence of SHP2 catalytic domain) which was used as the control experiment. The broken line indicates saturation transfer in the difference spectra containing both SHP2 catalytic domain and compound NSC-87877 (red), but not in the difference spectra recorded for compound NSC-87877 only (green).

### 4.2.4 Phosphatase activity assays in the presence of SHP2 inhibitors

As a final assessment of the activities of SHP2 inhibitors IIB08 and NSC-87877, analysis of inhibitory activity was investigated by examining the effect of compounds IIB08 and NSC-87877 on the inhibition of SHP2-catalysed hydrolysis of p-nitrophenyl phosphate (pNPP) using an in-vitro enzymatic assay. This provided a basis for confirming the biochemical activity of these compounds as well as observing whether the modes of inhibition corroborated with previously reported literature findings. Prior to evaluation of compounds IIB08 and NSC-87877 activity in the biochemical assay, preliminary experiments were conducted to assay the basal phosphatase activity of SHP2 catalytic domain (at a final concentration of 100 nM) in the absence of inhibitor allowing the derivation of a $K_m$ value for the pNPP substrate. Indeed, after several rounds of assay optimisation, values for the $K_m$ of the pNPP substrate for the SHP2 catalytic domain were calculated to be 5.1 mM (Figure 4.5A), 4.9 mM (in the presence of BSA; Figure 4.5B) and 5.2 mM (in the presence of Tween 20; Figure 4.5C) all of which are in close agreement with $K_m$ values ranging between 0.5-10 mM reported from independent kinetic experiments on multiple phosphatases [212-215].

To determine the mode of inhibition for compounds IIB08 and NSC87877, inhibitor assays were repeated in the presence of at least two fixed concentrations of each inhibitor ranging from 0.2-5-fold the apparent $K_m$ value of pNPP. Analysis of the activity data (Sigmaplot 12.0) enabled subsequent derivation of the Michaelis-Menten parameters $K_m$ and $V_{max}$ values. Consistently similar $K_m$ values of 5.1 mM, 5.0 mM and 5.3 mM recorded in the presence of inhibitor IIB08, along with a reduction in the $V_{max}$ values from 0.062 $\mu M/min^{-1}$, 0.051 $\mu M/min^{-1}$ and 0.038 $\mu M/min^{-1}$ at 0 μM, 5 μM and 10 μM

concentrations of IIB08 indicate a non-competitive mode of inhibition which corroborates with previously reported modes of inhibition for this compound using a similar assay [143]. In the case of inhibitor NSC-87877, a noticeable increase in the $K_m$ values from 5.1 μM, 8.6 μM, 11.4 μM and 13.7 μM were observed as a function of increasing inhibitor concentrations at 0 μM, 0.5 μM, 1.0 μM, 2.0 μM, which were also accompanied by a constant $V_{max}$ value of 0.062-0.065 μM/min$^{-1}$ indicating a competitive mode of inhibition. The observed mode of inhibitor action for NSC-87877 was also akin to previous observations reporting a similar mechanism of inhibition [136]. Furthermore, derivation of inhibitor constant ($K_i$) values (by fitting of the curves to a modified form of the Michaelis-Menten equation accounting for the presence of competitive and non-competitive inhibitor action) of 6.2 μM and 2.7 μM for the presence of compounds IIB08 (Figure 4.5A) and NSC-87877 (Figure 4.6A) respectively, were also in close agreement with the literature values of 5.5 μM and 0.3 μM previously reported for these compounds [136, 143].

To evaluate whether the inhibitory effects observed were a result of specific as opposed to non-specific inhibition of SHP2 catalytic domain, activity measurements were repeated using the same experimental parameters in the presence and absence of BSA (0.1 mg/ml), to observe whether this had any significant effect on the inhibition constant $K_i$. Analysis of the activity data for compounds IIB08 and NSC-87877 confirmed no significant difference between the inhibitor constants (6.2 μM for inhibitor IIB08 (Figure 4.5B), and 2.7 μM for inhibitor NSC-87877 (Figure 4.6B) when comparing measurements recorded in presence and absence of BSA, suggesting that inhibition of SHP2 phosphatase activity by inhibitors IIB08 and NSC-87877 is likely to be occurring through a specific rather than non-specific interaction. To address the possibility of aggregation-based mechanisms of inhibition which based on previous observations were suggested to underlie the behaviour of

compound IIB08 in the SPR assay, the inhibition assays for compounds IIB08 and NSC-87877 were repeated in the presence of the detergent Tween 20. As such, no significant difference in the inhibitor constant $K_i$ (5.9 µM for inhibitor IIB08 (Figure 4.5C) and 2.5 µM for inhibitor NSC-87877 (Figure 4.6C) was observed when compared with previous assays conducted under detergent-free conditions, suggesting that inhibition of SHP2-catalysed hydrolysis of pNPP is occurring through specific inhibition as opposed to inhibition driven by aggregation-based mechanisms.

A summary of the overall performance of inhibitors PHPS1 , IIB08 and NSC-87877 across the assay panel are summarised in Table 4.1, and the statistics from Sigmaplot analysis summarised in Table 4.2.

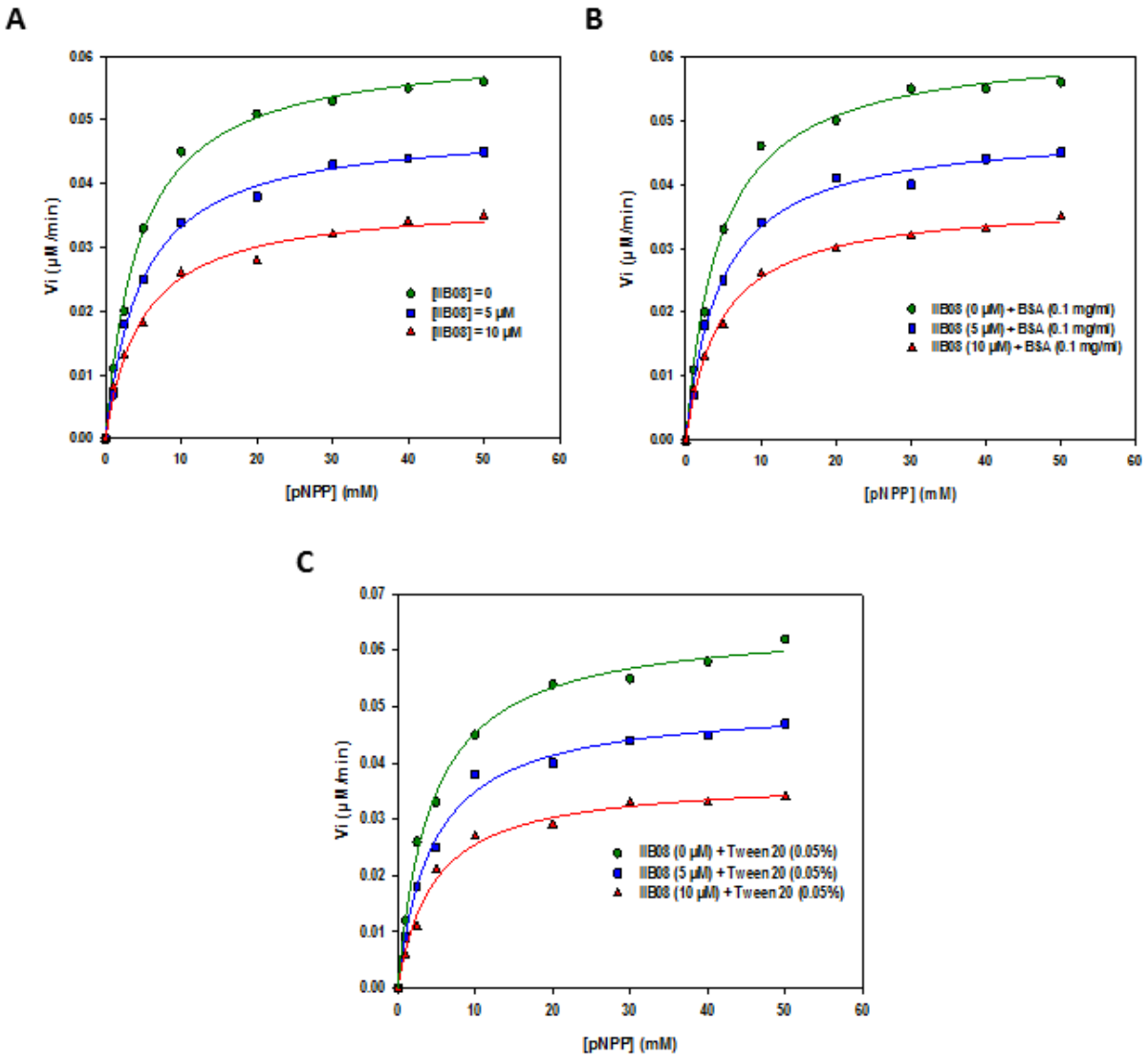



**Figure 4.5 Non-competitive inhibition of SHP2-catalysed pNPP hydrolysis by compound IIB08**. Inhibition of SHP2 catalytic activity was recorded at IIB08 concentrations of 0 µM, 5 µM and 10 µM at pNPP concentrations of 0-50 mM (0.5 $K_m$-10 $K_m$) and a final enzyme concentration of 100 nM. A) Inhibition of SHP2 catalytic activity in the presence 0 µM, 5 µM and 10 µM concentrations of compound IIB08 only. B) Inhibition of SHP2 catalytic activity with 0 µM, 5 µM and 10 µM concentrations of compound IIB08 in the presence of BSA (0.1 mg/ml). C) Inhibition of SHP2 catalytic activity with compound IIB08 in the presence of Tween-20 detergent (0.05%). All assays investigating the inhibitory activity of compound IIB08 were initiated with SHP2 catalytic domain at a final concentration of 100 nM.

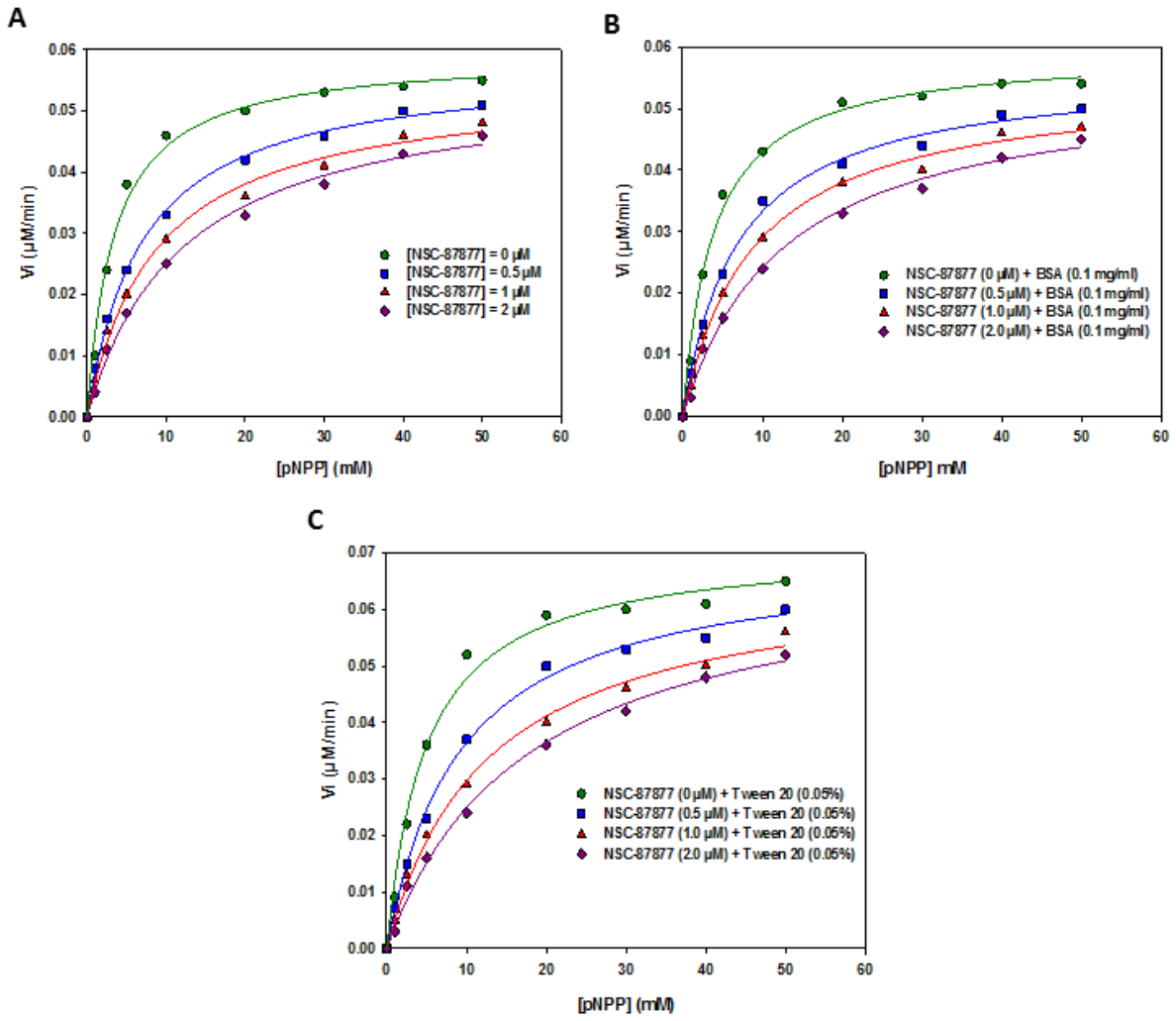



**Figure 4.6 Competitive inhibition of SHP2-catalysed pNPP hydrolysis by compound NSC-87877**. Inhibition of SHP2 catalytic activity was recorded at NSC-87877 concentrations of 0 µM, 0.5 µM, 1.0 µM and 2.0 µM at pNPP concentrations of 0-50 mM (0.5 $K_m$-10 $K_m$) and a final enzyme concentration of 100 nM. A) Inhibition of SHP2 catalytic activity with 0 µM, 0.5 µM, 1.0 µM and 2.0 µM concentrations of compound NSC-87877 only. B) Inhibition of SHP2 catalytic activity with 0 µM, 0.5 µM, 1.0 µM and 2.0 µM compound NSC-87877 in the presence of BSA (0.1 mg/ml). C) Inhibition of SHP2 catalytic activity with compound NSC-87877 in the presence of Tween-20 detergent (0.05%). All assays investigating the inhibitory activity of compound NSC-87877 were initiated with SHP2 catalytic domain at a final concentration of 100 nM.

Figure 4.5A statistics:

| | V..,.x(I.IM min·1) | K..,pNPP (mM) | K""'(s·1) | R2 | t-value | p-value |
|---|---|---|---|---|---|---|
| pNPPonly | 0.062 | 5.1 | 37.2 | 0.983 | 2.8 | 0.039 |
| 11808 (51.1M) | 0.051 | 5.0 | 30.6 | 0.981 | 2.6 | 0.041 |
| 11808 (10 !JM) | 0.038 | 5.3 | 22.8 | 0.989 | 2.9 | 0.038 |

Figure 4.58 statistics:

| | Vmax (IJM min·1) | KmpNPP (mM) | Kcat (s·1) | R2 | t-value | p-value |
|---|---|---|---|---|---|---|
| pNPP+ 8SA | 0.063 | 4.9 | 37.8 | 0.972 | 2.7 | 0.042 |
| 11808 (51JM) +8SA | 0.054 | 5.0 | 32.4 | 0.981 | 2.8 | 0.040 |
| 11808 (10 !JM) + 8SA | 0.041 | 5.1 | 24.6 | 0.962 | 2.7 | 0.038 |

Figure 4.5C statistics:

| | Vmax (IJM min·1) | KmpNPP (mM) | Kcat (s·1) | R2 | t-value | p-value |
|---|---|---|---|---|---|---|
| pNPP + Tween 20 | 0.068 | 5.2 | 40.8 | 0.964 | 2.1 | 0.035 |
| 11808 (5 !JM) +Tween 20 | 0.052 | 4.9 | 31.2 | 0.978 | 2.2 | 0.037 |
| 11808 (10 !JM) + Tween 20 | 0.039 | 5.1 | 23.4 | 0.983 | 2.4 | 0.031 |

Figure 4.6A statistics:

| | Vmax (liM min$^{-1}$) | KmpNPP (mM) | Kcat (s$^{-1}$) | R2 | t-value | p-value |
|---|---|---|---|---|---|---|
| pNPPonly | 0.065 | 5.1 | 39.0 | 0.962 | 2.3 | 0.041 |
| NSC-87877 (0.5 11M) | 0.064 | 8.6 | 38.4 | 0.984 | 2.6 | 0.031 |
| NSC-87877 (111M) | 0.062 | 11.4 | 37.2 | 0.953 | 2.7 | 0.036 |
| NSC-87877 (2 11M) | 0.062 | 13.7 | 37.2 | 0.626 | 2.2 | 0.030 |

Figure 4.68 statistics:

| | Vmax (liM min$^{-1}$) | KmpNPP (mM) | Kcat (s$^{-1}$) | R2 | t-value | p-value |
|---|---|---|---|---|---|---|
| pNPP+ BSA | 0.066 | 5.4 | 39.6 | 0.986 | 2.5 | 0.034 |
| NSC-87877 (0.5 11M) + BSA | 0.063 | 9.1 | 37.8 | 0.981 | 2.0 | 0.039 |
| NSC-87877 (111M) + BSA | 0.063 | 12.2 | 37.8 | 0.945 | 2.2 | 0.032 |
| NSC-87877 (2 11M) + BSA | 0.061 | 14.7 | 36.6 | 0.974 | 2.4 | 0.037 |

Figure 4.6C statistics:

| | Vmax (liM min$^{-1}$) | KmpNPP (mM) | Kcat (s$^{-1}$) | R2 | t-value | p-value |
|---|---|---|---|---|---|---|
| pNPP + Tween 20 | 0.072 | 5.6 | 43.2 | 0.973 | 2.9 | 0.028 |
| NSC-87877 (0.5 11M)+ Tween20 | 0.069 | 8.8 | 41.4 | 0.979 | 2.4 | 0.032 |
| NSC-87877 (111M) + Tween20 | 0.068 | 12.4 | 40.8 | 0.981 | 2.6 | 0.035 |
| NSC-87877 (2 11M) + Tween20 | 0.068 | 15.1 | 40.8 | 0.969 | 2.1 | 0.039 |

| SHP2 inhibitor | Apparent $K_i$ (μM) | Expected $K_i$ (μM) |
|---|---|---|
| IIB08 | 6.2 | 5.5 |
| IIB08 + BSA | 6.0 | 5.5 |
| IIB08 + Tween 20 | 5.9 | 5.5 |
| | | |
| NSC-87877 | 2.7 | 0.3 |
| NSC-87877 + BSA | 2.8 | 0.3 |
| NSC-87877 + Tween 20 | 2.5 | 0.3 |

**Table 4.1 Determination of Michaelis-Menten and statistical parameters from non-linear regression analysis of SHP2 catalytic activity in the presence and absence of inhibitors IIB08 and NSC-87877.** Enzyme kinetic and statistical parameters for SHP2 activity were determined from individual curves showing SHP2 activity in the presence of pNPP as well as in the presence of inhibitors IIB08 and NSC-87877. These parameters were also determined for repeats of these experiments conducted in the presence and absence of BSA and Tween 20. The apparent values for the inhibitor constants ($K_i$) for IIB08 and NSC-87877 have also been derived and then compared with the expected values previously reported in the scientific literature. $V_{max}$ values for individual experiments indicate the maximum velocity of the enzyme-catalysed reaction occurring at saturating concentrations of substrate. The $K_m$ values represent the concentration of substrate required to give half maximum velocity ($1/2V_{max}$) of the enzyme-catalysed reaction. The $K_{cat}$ of an enzyme is often referred to as the turnover number and is a measurement of the catalytic efficiency of the enzyme which signifies the number of substrate molecules turned over into product per enzyme per second. The inhibitory constant or Ki is the inhibitor concentration required to produce half maximum inhibition and serves as a measure of inhibitor potency. Statistical parameters derived from non-linear regression fitting of the individual curves in Sigmaplot have been provided for the highest concentration of inhibitors IIB08 and NSC87877 measured in the presence and absence of BSA and Tween 20. The $R^2$ value represents the coefficient of determination; the closer the values of $R^2$ to 1 the greater the accuracy of the fitted curve. The t value corresponds to how accurately the parameters (e.g. $V_{max}$ and $K_m$) predict the values of the reaction velocity ($V_o$); the larger the t value the more the value of $V_o$ depends upon the parameters of $K_m$ and $V_{max}$. The p-value represents the probability of the $K_m$ and Vmax values incorrectly predicting the values of $V_o$. The p value should be <0.05.

| | | $K_D$ (µM) | | | $K_i$ (µM) |
|---|---|---|---|---|---|
| **Compound** | 2D-NMR | SPR | STD-NMR | waterLOGSY | Enzyme assay |
| **PHPS1 (1)** | No binding | No binding | No binding | No binding | Not tested |
| **IIB08 (2)** | No binding | Ambiguous | No binding | No binding | 3 µM |
| **NSC-87877 (3)** | Binding* | 2 µM | Binding | Binding | 0.5 µM |

**Table 4.2. Summary of biophysical and biochemical assay profiling of SHP2 inhibitor compounds PHPS1, IIB08 and NSC-87877**. Compound PHPS1 was found to be highly insoluble in 5% (v/v) DMSO and thus attempts to detect an interaction with the SHP2 catalytic domain by NMR and SPR proved unsuccessful. Although more soluble than compound PHPS1, IIB08 also demonstrated poor solubility at concentrations above 50 µM in 5% (v/v) DMSO. As such, interactions between IIB08 and SHP2 catalytic domain were not observed by NMR when tested in both STD and waterLOGSY experiments. Although initially promising, SPR characterisation of compound IIB08 interaction with SHP2 catalytic domain was later suggested to be non-specific due to aggregation at concentration of 50 µM and above. Compound NSC-87877 exhibited consistent activity across the panel of assays and thus was used as a validated positive control to authenticate future screening assays.

## 4.3 Discussion

As a precursor to fragment-based ligand discovery to accelerate the development of more selective and potent SHP2 chemical probes, a series of robust biophysical and enzyme inhibition assays were developed and optimised to examine the binding affinities and biochemical activities of three previously reported SHP2 inhibitors, PHPS1, IIB08 and NSC-87877. Although tested in preliminary screening measurements, we decided to prematurely exclude inhibitor PHPS1 from further studies midway through our investigation owing to a profound lack of compound solubility. As such, this compound was deemed unsuitable for use as an assay-validating chemical tool.

Prior to engaging in fragment-based screening of SHP2, it was deemed necessary to invest efforts in profiling the activities of previously reported SHP2 chemical probes across our panel of relevant screening assays to determine the utility of these probes as potential tool compounds for hit validation in fragment screening. The cascade of small-molecule screening assays included protein and ligand-observed NMR spectroscopy, SPR, and enzymatic activity assays, all of which enabled determination and confirmation of binding interactions, affinities and inhibitory mechanisms of action.

Compounds NSC-87877 and IIB08 were initially selected for investigation on the basis of their literature precedent as validated and commercially available SHP2 inhibitors [136, 143]. The initial identification of compound NSC-87877 emerged through screening of the National Cancer Institute (NCI) Diversity chemical library and validation of inhibitory activity was confirmed in enzymatic pNPP assays with a reported $IC_{50}$ of 0.3 μM against SHP2 along with cross-inhibition of the closest SHP2 homologue, SHP1, with a similar potency. In addition to characterising inhibition using biochemical assays, the deployment

of molecular docking analysis in combination with mutagenesis studies suggested an active-site directed mode of SHP2 inhibition, validation of which has yet to be confirmed experimentally [136]. Conversely, compound IIB08 was identified by employing a unique combinatorial library approach which utilised an indole salicylic acid core as an initial chemotype to accelerate the search for novel bivalent inhibitors of SHP2. In a pNPP biochemical assay, compound IIB08 gave an $IC_{50}$ of 5.5 μM and a modest threefold selectivity profile over close phosphatase homologs SHP1 and PTP1B. Subsequent kinetic analysis also revealed compound IIB08 exhibited reversible non-competitive inhibitor kinetics. These kinetic observations were accompanied by promising inhibitory activity in a cell-based assay where compound IIB08 demonstrated effective repression of EGF-triggered activation of ERK1/2 leading to sustained hyper- proliferation of haematopoietic progenitors in response to GM-CSF cytokine signalling [143]. Observations of the inhibitory effects of compound IIB08 have also successfully translated into in-vivo studies resulting in abrogation of xenograft tumour growth in mouse models of non-small cell lung carcinoma [142]. Despite compelling evidence for the inhibitory potential of these compounds in enzymatic and cell-based assays, the biophysical evidence confirming binding modes, affinities and kinetic parameters for inhibitor interactions with the SHP2 catalytic domain remains surprisingly underexplored. To address this imbalance, a range of biophysical assays were deployed to directly characterise inhibitor interactions in an attempt to validate compounds IIB08 and NSC-87877 as authentic and biophysically-active chemical tools which could be harnessed as positive control compounds during fragment-based screening of SHP2.

Interrogation of binding interactions between compounds IIB08 and NSC-87877 and SHP2 catalytic domain were first conducted by $^{15}N$-$^{1}H$ HSQC NMR spectroscopy. Despite

the unique sensitivity of this method for mapping inhibitor binding sites through perturbations in backbone amide signals, the poor solubility of IIB08 at the high concentrations required for inhibitor binding obfuscated any attempts to observe inhibitor binding. Although titration of inhibitor NSC-87877 against $^{15}$N-labelled SHP2 catalytic domain yielded an absence of observable chemical shift perturbations, putative residues involved in this interaction (described in Chapter 3) were already identified providing support for a pre-existing in-silico model of NSC-87877 binding [136].

In conjunction with protein-observed NMR methods, the biophysical assay cascade also exploited the use of SPR to provide orthogonal evaluation and direct observation of compound IIB08 and NSC-87877 binding to the SHP2 catalytic domain. This enabled derivation and confirmation of $K_D$ values of 2 µM for compound NSC-87877 and 3 µM for IIB08 interactions with the SHP2 catalytic domain. Inspection of the raw SPR data for compound NSC-87877 binding to SHP2 catalytic domain revealed a binding profile indicative of a specific and reversible binding interaction. However, although initial observations suggested a specific interaction between compound IIB08 and SHP2 catalytic domain, closer analysis of the response versus time data (Figure 4.3B) suggested that the binding profiles observed may not be truly representative of a specific binding interaction. This is supported by the absence of true saturation of the response signal as a function of increasing inhibitor concentration. Instead, the change in response signal was observed to increase by the same increment each time as a function of the inhibitor concentration indicating an aggregation-based mechanism of interaction. This suggests that compound IIB08 may be acting in a bi-phasic manner where, initially, a specific interaction is observed between compound IIB08 and the catalytic domain of SHP2 at compound concentrations below 50 µM. However, as the concentration of IIB08 exceeds 100 µM, the

SPR curve displays a linear increase in response characteristic of an aggregation-driven interaction. The use of IIB08 as a tool compound is severely compromised by its limited solubility and any results of its inhibitory properties must therefore be treated with caution.

To finalise the characterisation of compound IIB08 and NSC-87877, inhibitor constants were determined for each inhibitor using enzymatic assay based on the spectrophotometric detection of p-nitrophenol production as a result of SHP2-catalysed hydrolysis of pNPP. In this assay, compounds IIB08 and NSC-87877 gave $K_i$ values of 6.2 µM and 2.8 µM respectively, both of which are in close general agreement with the previously reported literature values of 5.5 µM and 3 µM for these compounds. Analysis of the inhibitory mechanism revealed that compounds IIB08 and NSC-87877 demonstrate non-competitive and competitive inhibitor kinetics respectively, which corroborate with previously reported modes of inhibition for these compounds [136, 143]. Furthermore, repeating the inhibition assays for compounds IIB08 and NSC-87877 in the presence of BSA (0.1 mg/ml) revealed no significant difference in kinetic parameters for SHP2 activity compared to inhibitor assays conducted in the absence of BSA suggesting that inhibition of SHP2-catalysed hydrolysis of pNPP occurred as a result of specific inhibition as opposed to non-specific promiscuous inhibition of BSA. This confirmed, at least mechanistically, that compounds IIB08 and NSC-87877 appeared to be operating through specific rather than non-specific inhibitor mechanisms of action in the pNPP biochemical assay. Interestingly, analysis of compound IIB08 behaviour across the spectrum of binding assays not only highlighted issues relating to compound solubility, but also raised concerns regarding the possibility of aggregation-based mechanisms of inhibition, which was postulated as a potential mechanism underlying the behaviour of this compound in the SPR assay. It has been well documented that at micromolar concentrations, many small molecules have the propensity

to self-assemble into micelle-like aggregates capable of driving promiscuous and non-specific inhibition of enzymes and proteins [152, 216]. Attenuation of inhibition under detergent-rich conditions is therefore a key hallmark of aggregation-based mechanisms of inhibition [217, 218]. As such, repeating the inhibition assays for compounds IIB08 and NSC-87877 in the presence of the detergent Tween 20 confirmed that there was no significant difference in the values derived for the inhibitor constant $K_i$ compared with previous assays conducted under detergent-free conditions, suggesting that inhibition of SHP2-catalysed hydrolysis of pNPP is occurring through specific inhibition as opposed to inhibition driven by aggregation-based mechanisms.

Despite extensive characterisation of NSC-87877 activity in enzymatic assays and cell culture reported in previous studies [219] the availability of structural evidence in support of a competitive mode of inhibition remains elusive. In contrast, the previously reported co-crystal structure of IIB08 in complex with SHP2 catalytic domain revealed that IIB08 binds preferentially to the phosphatase active site supporting a competitive inhibitor mechanism of action.

## 4.4 Conclusions

The binding interactions and biochemical activities of previously reported SHP2 inhibitors IIB08 and NSC-87877 were characterised and their performance compared across a cascade of relevant and robust biophysical and biochemical assays, including 2D-NMR, STD-NMR, waterLOGSY, SPR and enzymatic assays. These observations have confirmed that inhibitor NSC-87877 demonstrates consistent activity across all assays, and appears to be a validated SHP2 inhibitor which can be exploited for use as a positive control compound in fragment-based screening. Although attempts to characterise compound

IIB08 as a non-competitive inhibitor of SHP2 were successful in the biochemical assay, the observation of an active-site binding mode from a previously determined co-crystal structure of this compound in complex with the SHP2 catalytic domain is intriguing, especially given the absence of inhibitor contacts outside the active site cavity. Overall therefore, it is still unclear whether the observed effects are a result of true inhibition or a false positive driven by compound aggregation in the NMR and SPR assays. Therefore, the inability of IIB08 to manifest activity consistently across the NMR and SPR assays suggest that, for the purposes of this investigation, it was intractable for use as a validated chemical probe.

# CHAPTER 5: FRAGMENT-BASED SCREENING OF SHP2 BY NMR AND SPR

## 5.1 Experimental aims

In Chapter 3, protein-observed NMR spectroscopy was employed to interrogate ligand binding to the SHP2 catalytic domain under physiologically-relevant conditions, yielding novel insights into the structural basis of ligand interactions in the solution. In Chapter 4, a range of biophysical and biochemical assays were conducted in order to validate the binding affinities and biochemical activities of previously reported SHP2 inhibitors, providing a robust evaluation of the suitability of these ligands for use as positive controls in fragment-based screening of SHP2. The utility of validated chemical tools with demonstrable binding interactions was a necessary prerequisite for enabling validation of our screening approach and ensured that the SPR and NMR-based assays used for fragment screening were sufficiently configured for optimal hit detection.

The over-arching aim of this chapter was to utilise SPR and NMR-based approaches to screen a 433-compound fragment library with a diverse range of chemical fragments (Figure 5.1; see Appendix figures 3a and 3b for extended table highlighting the chemical identities of all 433 fragments) against the SHP2 catalytic domain, Tandem-SH2 domains and full length SHP2, in order to identify fragment hits that could serve as chemical starting points for the development of more potent SHP2 inhibitors.

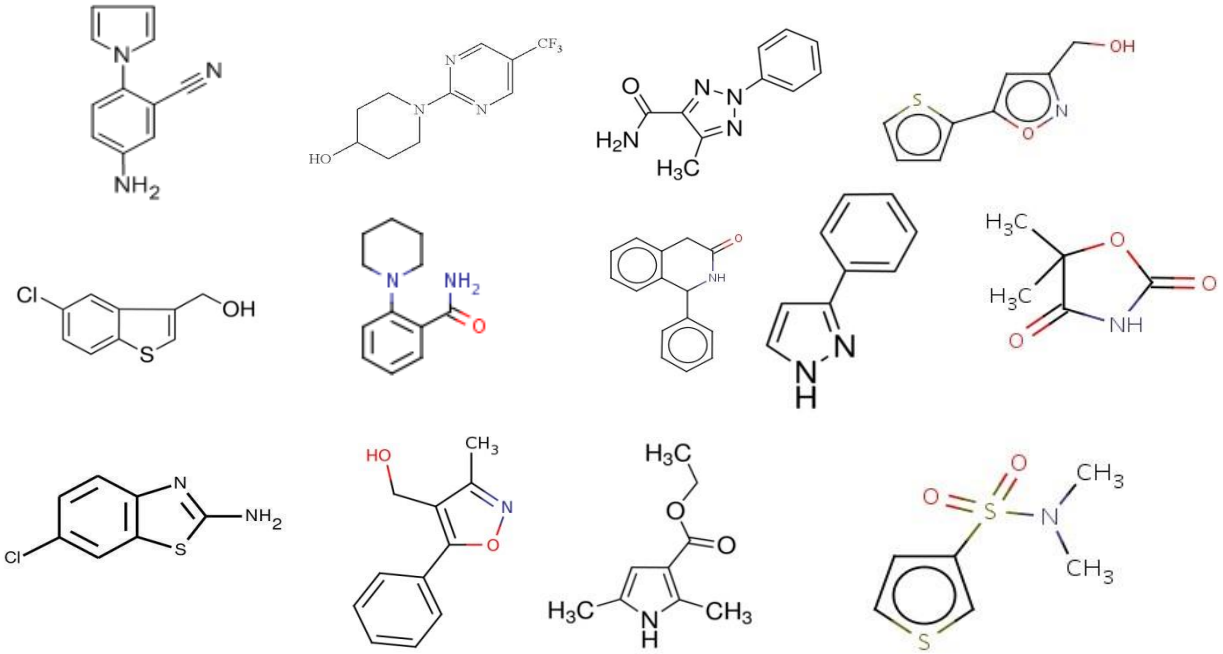


**Figure 5.1 Chemical composition of fragment library.** A 433-compound fragment library was assembled from a mixture of fragments deriving from the Maybridge and Sigma fragment collections. The fragment scaffolds were selected according to a checklist of pre-determined selection critieria. Such criteria included physical properties, such as low molecular weight (110 Da - 250 Da) and solubility in water (2 mM), chemical tractability ensuring fragment contain appropriate functional groupss for synthesis, diversity of chemistry, optimal purity and stability, and applicability to a range of target classes.

## 5.2 Results

Prior to implementing a fragment-based screening strategy to accelerate the development of novel SHP2 inhibitors, a 433-compound fragment library of sufficient chemical diversity (Figure 5.1) was compiled from a selection of commercially available chemical fragments from the Maybridge and Sigma collection. Individual fragments were selected on the basis of physiochemical properties including molecular weight (typically between 110 Da - 250 Da), predicted solubility in water (2 mM), suitability for the proposed screening strategy, chemical tractability, appropriate functionality for synthesis, diversity of chemical scaffolds and general applicability to a wide array of target classes. All compounds were dissolved in 100% $D_6$-DMSO to generate stock solutions at a final concentration of 200 mM. Compounds were diluted to 1 mM in 5% $D_6$-DMSO and were subjected to $^1$H-NMR quality control analysis before screening to confirm structural integrity and verify chemical identity. Mass spectrometry on selected compounds was also performed to resolve any ambiguities in compound validation. The preparation of the fragment stocks and quality control measures were conducted by previous project students Leanne Davies and Mussa Qureshy prior to starting this PhD project.

### 5.2.1 Expression and purification SHP2 expression constructs for SPR

In order to facilitate fragment-library screening of SHP2 using SPR, a series of expression constructs comprising SHP2 catalytic domain (residues 247-521) (Figure 5.2), SHP2 Tandem SH2 domain (residues 6-216) (Figure 5.3), and wild type full-length SHP2 + C-terminal tail (residues 1-598) (Figure 5.4) were designed and engineered to contain a C-terminal Avi-tag™ sequence (GLNDIFEAQKIEWHE) to aid in-vitro biotinylation of SHP2 for SPR immobilisation. All constructs were purposely engineered with the Avi-

tag™ at the C-terminus in order to avoid the N-terminal $His_{x6}$-tag and to preserve the native dynamics and allosteric inhibitory mechanisms in the full length construct that might otherwise be compromised by immobilisation via the N-terminus. All SPR Avi-tag™ expression constructs were codon-optimised for heterologous expression in *E. coli*, cloned into appropriate expression vectors and expressed and purified as described in Materials and methods (see section 2.2.6).

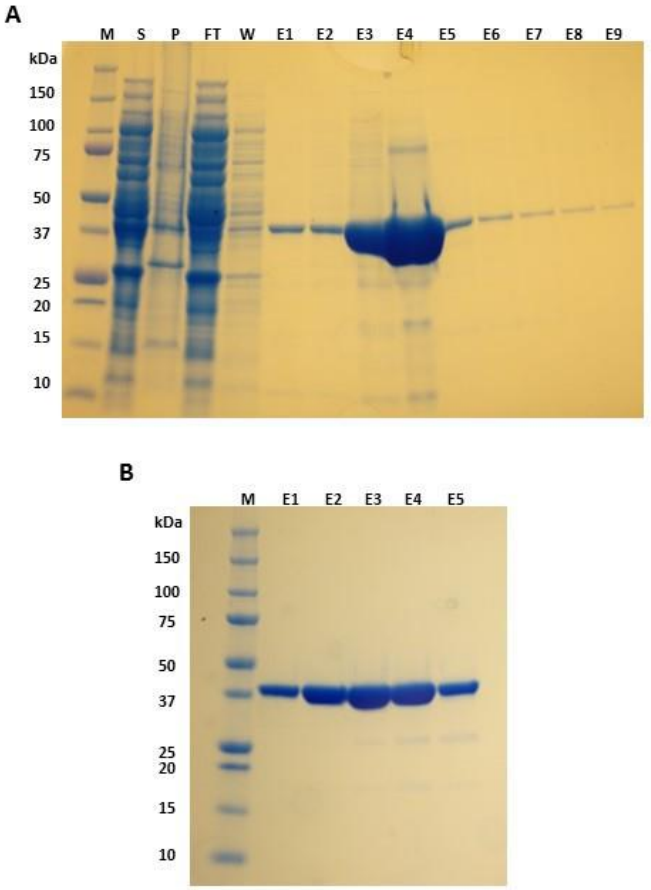



**Figure 5.2. Expression and purification of the SHP2 catalytic domain Avi-tag™ construct.** A. Purification of SHP2 catalytic domain by $Ni^{2+}$-NTA affinity chromatography. From left to right: molecular weight markers (M), supernatant (S), flowthrough (FT), wash (W), elution fractions (E1-E9). B. Purification of SHP2 catalytic domain by size-exclusion chromatography. From left to right: molecular weight markers (M), elution fractions (E1-E5).

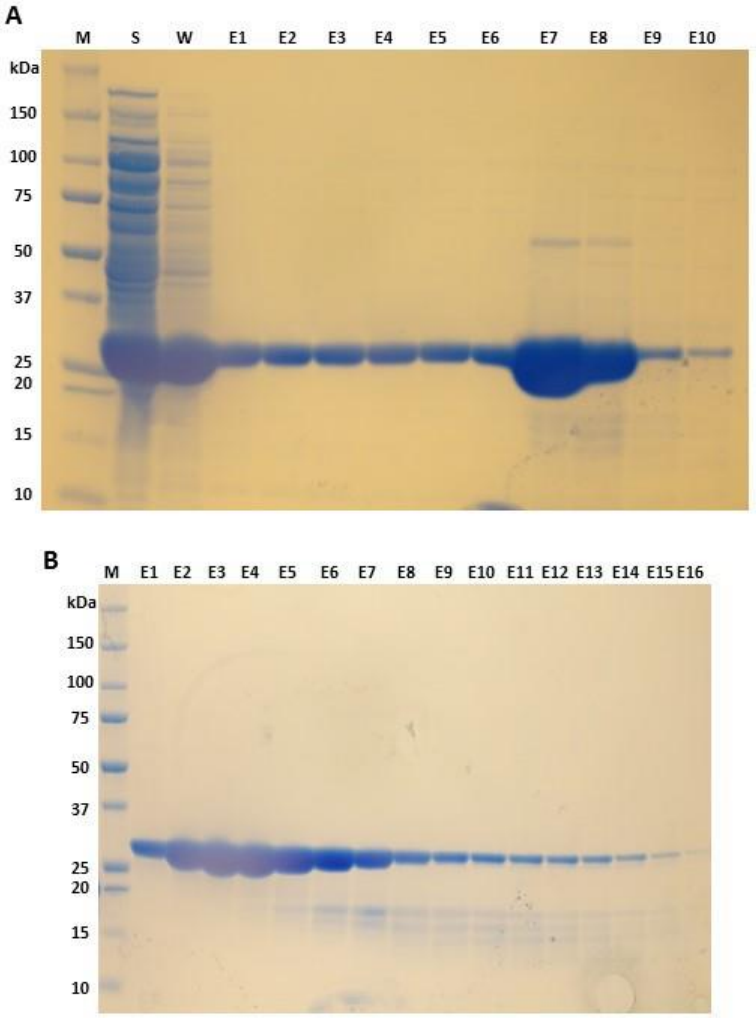



**Figure 5.3. Expression and purification of SHP2 Tan-SH2 domain Avi-tag™ construct.** A. Purification of SHP2 Tan-SH2 domain by $Ni^{2+}$-NTA affinity chromatography. From left to right: molecular weight markers (M), supernatant (S), wash (W), elution fractions (E1-E10). B. Purification of SHP2 Tan-SH2 domain by size-exclusion chromatography. From left to right: molecular weight markers (M), elution fractions (E1-E16).

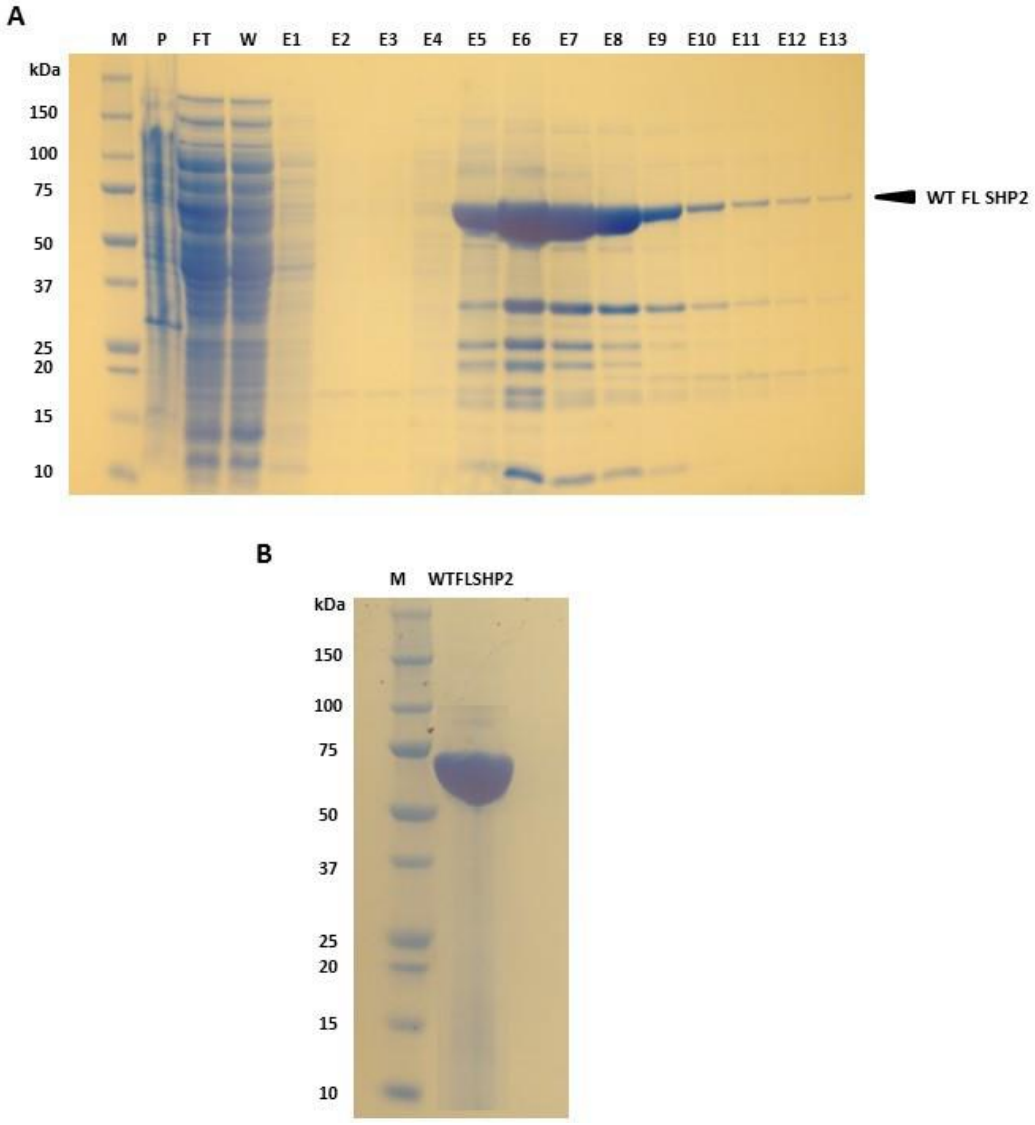



**Figure 5.4. Expression and purification of full length SHP2 Avi-tag™ construct.** A. Purification of full length SHP2 Avi-tag™ construct by $Ni^{2+}$-NTA affinity chromatography. Arrowhead indicates full length SHP2 eluting from the $Ni^{2+}$-NTA column wit Ah molecular weight of approximately 68 KDa. From left to right: molecular weight markers (M), pellet (P), flowthrough (FT), wash (W), elution fractions (E1-E13). B. Purification of full length SHP2 by size-exclusion chromatography. From left to right: molecular weight markers (M), pooled and concentrated elution fractions of pure full length SHP2 from size exclusion chromatography.

### 5.2.2 Biotinylation of SHP2 catalytic, Tan-SH2 and full length SHP2 constructs

As a precursor to initiating fragment-screening of SHP2 catalytic, Tan-SH2 and full length SHP2 by SPR, an appropriate strategy was devised for generating biotinylated forms of each these constructs, in order to achieve protein immobilisation to the Biacore sensor chip surface via biotinylation of the C-terminal Avi-tag$^{TM}$. Preliminary experiments suggested that a single screening run of the entire 433-compound fragment library (with all fragments run as singletons) would take approximately three days for completion. Thus, adopting a strategy using biotinylated protein coupled to a streptavidin coated sensor chip surface ensured optimum stability and protein immobilisation throughout the entire duration of each individual screening run. As a precursor to protein immobilisation to the SPR sensor chip surface, each SHP2 construct was incubated with BirA ligase from *E coli,* which catalyses the ATP-dependant covalent attachment of free biotin [220] to lysine residues present within the biotin recognition sequence GLNDIFEAQKIEWHE. After experimenting with a range of biotin concentrations and incubation times, it was observed that incubation of the biotinylation reaction for 2 hours at a constant temperature of 30$^{o}$C along with a two-fold molar excess of biotin relative to protein concentration was sufficient to achieve detectable biotinylation of all SHP2 Avi-tag$^{TM}$ constructs. Throughout all experiments both temperature and buffer conditions (supplemented with $MgCl_2$ to provide an essential source of $Mg^{2+}$ as a cofactor for aiding BirA catalysis) were kept constant. After employing size-exclusion chromatography to purify biotinylated protein and facilitate the removal of excess biotin, western blotting analysis was used to confirm the presence of biotinylated species (Figure 5.5).

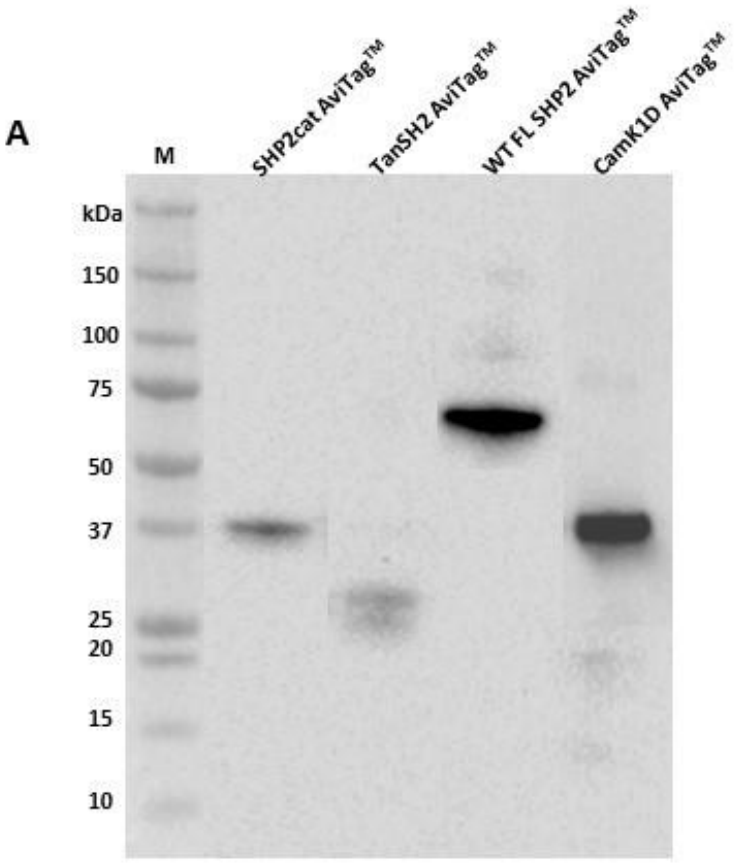



**Figure 5.5. Western blot analysis of biotinylated SHP2 catalytic domain, Tan-SH2 domain and full length SHP2**. Confirmation of successful biotinylation of SHP2-Avi-tag™ constructs was judged relative to biotinylated CamK1D control protein. From left to right: molecular weight markers (M), biotinylated SHP2 catalytic domain-Avi-tag™, biotinylated Tan-SH2 domain-Avi-tag™, biotinylated full-length SHP2-Avi-tag™ and biotinylated CamK1D-Avi-tag™ (positive control).

Although precise quantification of biotinylation levels for each species was not measured, successful biotinylation was determined by comparison against a control sample of biotinylated CamK1D protein (molecular weight = 38 KDa), which revealed successful but differential levels of biotinylation for the SHP2-Avi-tag™ construct. Since all reaction conditions were conducted using the same incubation times, ATP and biotin concentrations, the observation of differential levels of biotinylation could be rationalised by the relative accessibility of the C-terminal Avi-tag™ sequence for ligation by the BirA ligase enzyme. This was particularly pertinent for the full length SHP2 Avi-tag™ construct which resulted in the highest level of biotinylation relative to the control protein CamK1D. This was suspected to have arisen due to location of the Avi-tag™ which, being tethered to the end of the disordered SHP2 C-terminal tail would provide greater accessibility to BirA enzyme in solution to facilitate the biotinylation reaction. The same justification can also be applied to both the SHP2 catalytic and Tan-SH2-Avi-tag™ constructs, each of which showed comparatively less biotinylation compared with the full length SHP2 Avi-tag™ construct. For similar reasons, it was suggested that the lack of accessibility of the Avi-tag™ as a result of being tethered to the more globular SHP2 catalytic and tan-SH2 domains, may have sterically hindered BirA ligase recognition of the Avi-tag™ sequence, leading to less efficient biotinylation. After confirming successful biotinylation, each SHP2-Avi-tag™ construct was immobilised to the streptavidin-coated SPR sensor chip surface following several rounds of optimisation.

The optimum immobilisation level for each SHP2 construct for fragment screening was decided based on a molecular weight of 150 Da, which represents the average molecular weight calculated for all 433 compounds present in the fragment library. The average molecular weight (analyte MW) was then used in conjunction with the molecular weight of each SHP2 construct (ligand MW) to calculate the maximum theoretical response value (Rmax) based on a protein immobilisation level of 8,000 RU for each construct (Equation 1). For a small molecule fragment (molecular weight - 150 Da) binding to a protein (molecular weight – 34,000 Da) immobilised to a level yielding 8,000 RU, the maximum expected response would be 35 RU.

Equation 1:

$$R\text{[illegible]}(R\text{[illegible]}) = \frac{(\text{[illegible]})}{M\text{[illegible]}} \times i\text{[illegible]}(R\text{[illegible]})$$

The final protein concentrations were 1 µM, 1.5 µM and 3 µM for SHP2 catalytic domain, Tan-SH2 domain and full length SHP2 which each gave immobilisation levels of 7900 RU, 7875 RU and 7600 RU, respectively, compared to the response expected response of 8000 RU for each SHP2 construct.

### 5.2.3 Fragment-library screening by SPR

After confirming successful biotinylation levels for each SHP2 construct and ensuring sufficient levels of biotinylated protein had been immobilised to the streptavidin-coated sensor chip surface, a library of 433-fragments from the Maybridge-Sigma selection (see 2.8.1 and Appendix figures 3a and 3b) was screened by SPR to order to identify active

fragments that bound to SHP2 catalytic domain, Tan-SH2 domain and wild-type full length SHP2. The buffer screening conditions were devised based on earlier findings (as described in chapters 3 and 4) demonstrating that the SHP2 catalytic domain exists as a stable, monomeric and appropriately folded conformation in the presence of both phosphate and HEPES buffer conditions. These observations suggested that implementing a screening strategy employing both HEPES and phosphate buffer conditions would allow screening of SHP2 under two separate conditions both of which support structural integrity and functional phosphatase activity despite differing in their physiological relevance. Incorporating both screening conditions would also provide an interesting comparative dataset for assessing the difference in fragment hit rates between phosphate and non-phosphate buffer conditions, providing informative insights into future screening strategies for an emerging target class that remain largely unexplored for their drug discovery potential.

Each SPR fragment screen was completed in five separate experimental runs lasting a total of three days. The entire 433 fragments were screened at single concentrations of 50 µM and 200 µM in both HEPES and phosphate buffer conditions, with each screening run lasting for a total of three days. Each screen was conducted against C-terminally biotinylated SHP2 catalytic, Tan-SH2 and full length SHP2 Avi-tag$^{TM}$ constructs immobilised on a streptavidin-coated Series SA Biacore sensor chip to levels of ~7900, 7875 and 7600 RUs respectively. In order to assess the stability of SHP2 catalytic domain throughout the duration of the screening assay, the positive control compound NSC-87877 was injected at a concentration of 100 µM. Unfortunately, there were no small molecule inhibitors available for probing the stability of Tan-SH2 domains throughout the SPR screening assay. The same was also true for wild-type full length SHP2, however the

recent discovery of allosteric inhibitor SHP099 as the first reported inhibitor of full length SHP2 could have been exploited for use as a stability probe in these assays had earlier development been possible [149] .

The data from all SPR screening runs were processed as described in the Material and Methods (see section 2.8.4). Deconvolution of fragment hit matter was determined by setting a cut-off level of 3 RU according to previously published methods [221]. The sensorgram profiles for suspected hits at or above this cut-off were then interrogated individually for overall shape to identify specific binders and eliminate non-specific fragment hits (Figure 5.6).

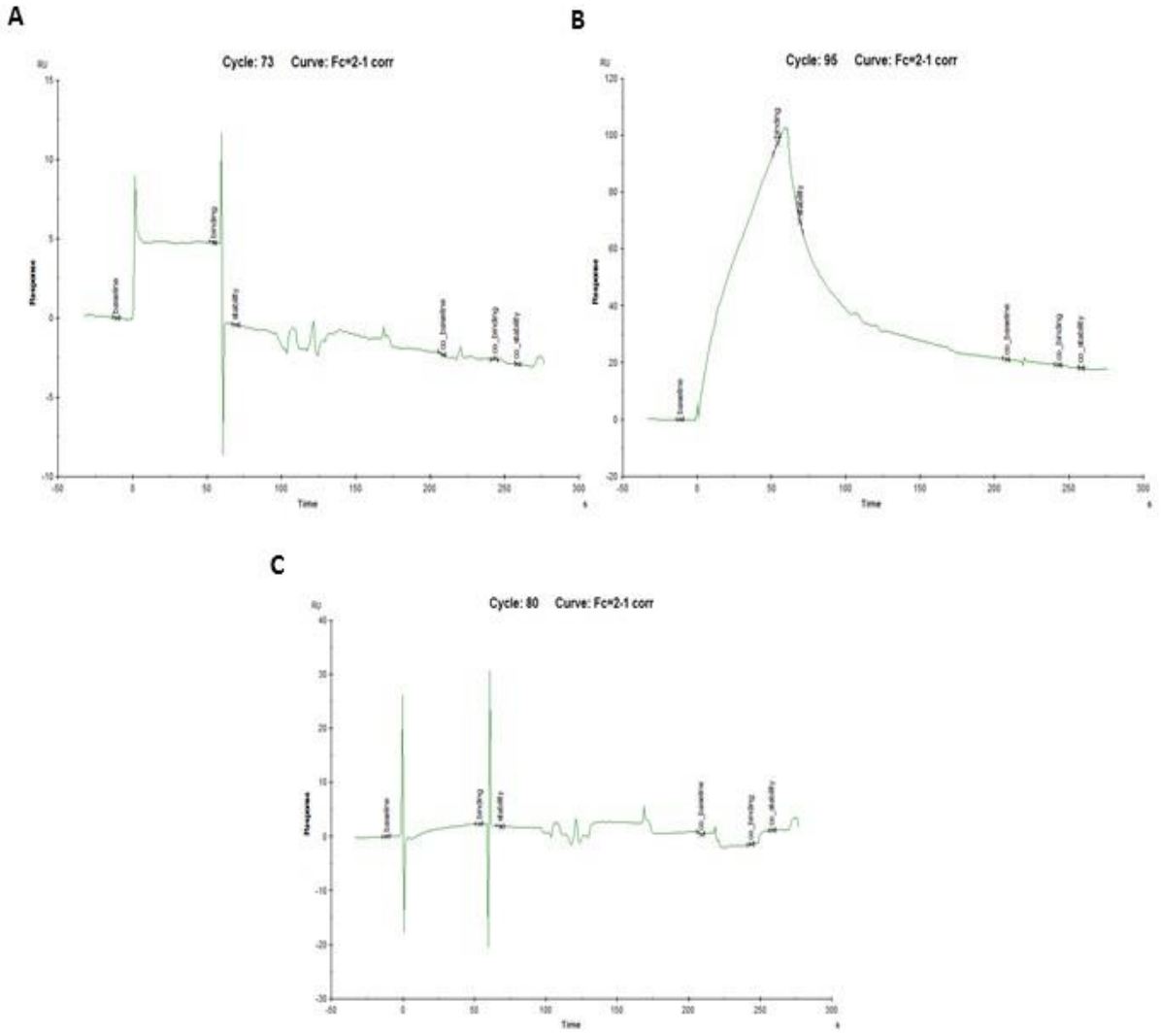



**Figure 5.6. Interrogation of SPR sensorgram profiles from fragment-based screening of the SHP2 catalytic domain to discriminate between specific versus non-specific and non-binding fragments.** A) Classic "table-top" profile of a fragment hit demonstrating fast on/off kinetics indicative of specific fragment binding. B) Super-stoichiometric binding phenomena indicative of non-specific fragment binding. C) Sensorgram profile of a non-binding fragment showing negligible change in response.

### 5.2.4 Identification of active fragment hits against SHP2 catalytic domain by SPR

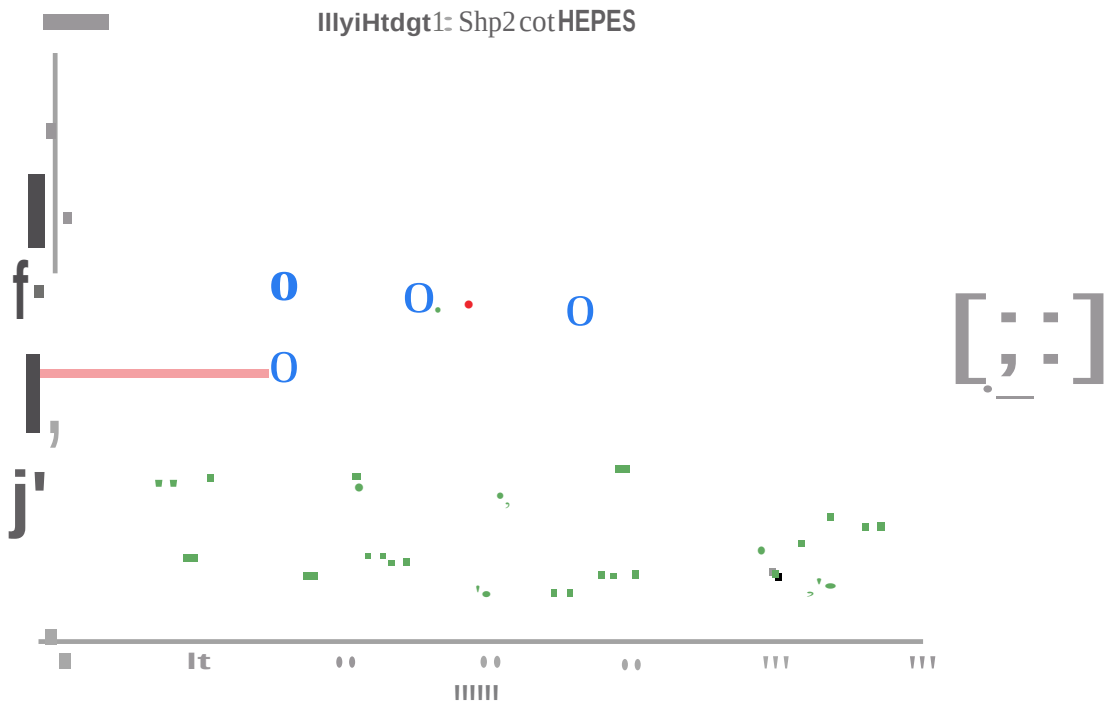
Unfortunately, preliminary attempts to identify fragment hits against SHP2 catalytic, tan-SH2 and full length SHP2 domains when screening at a single fragment concentration of 50 µM in both phosphate and HEPES buffer conditions (data not shown) were unsuccessful. Owing to the lack of fragment hits at 50 µM concentration, the fragment screening concentration was increased to 200 µM to increase the possibility of identifying active fragment hits against each domain. Indeed, screening at a single fragment concentration of 200 µM enabled identification of a total of 11 active fragments against the SHP2 catalytic domain in HEPES buffer (Figure 5.7), with each hit confirmed through rigorous inspection of individual sensorgram profiles. Interestingly, a repeat of the same 433 fragment screen in phosphate buffer revealed 6 active fragments against the SHP2 catalytic domain (data not shown), 2 of which were also identified as part of the 11 fragment hits originally identified in the HEPES buffer screen. Despite increasing the fragment screening concentration to 200 µM, identification of fragment hits against the Tan-SH2 domains in both HEPES and phosphate buffer conditions was still unsuccessful, exemplifying the challenges associated with the development of chemical tools for selective modulation of SH2 domain function. Equally, a similar but unexpectedly low fragment hit rate was also observed for the full length SHP2 domain, with only 5 fragment hits recorded in phosphate and none in HEPES buffer conditions when screening at a single fragment concentration of 200 µM (data not shown). This suggested that future assay protocols for SPR screening of Tan-SH2 and wild type full length SHP2 domains may benefit from screening single fragment concentrations at or exceeding 0.5 mM. Results showing the number of fragment hits recorded against SHP2 catalytic domain,

Tan-SH2 and full length SHP2 in each buffer screening condition are summarised in Table 5.1.

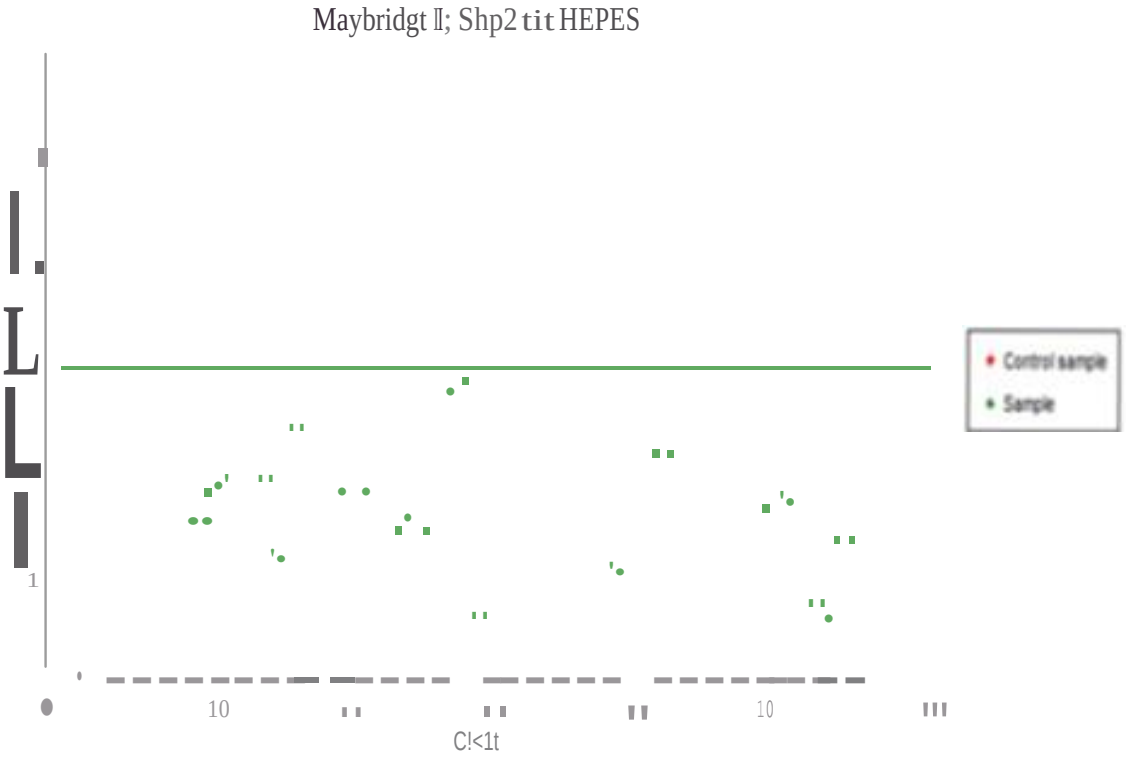
| | SPR | | | STD | | | waterLOGSY | | |
|---|---|---|---|---|---|---|---|---|---|
| SHP2construct | HEPES | NaP | Both | d-Tris | NaP | Both | d-Tris | NaP | Both |
| SHP2 cat | 11 | 6 | 2* | 27 | 14 | 2* | 24 | 14 | 2* |
| Tan-SH2 | 2 | 2 | 0 | 9 | 10 | 0 | 13 | 15 | 0 |
| FL SHP2 | 0 | 5 | 0 | 3 | 4 | 0 | 6 | 8 | 0 |
| SHP2 cat + Tan-SH2 | 13 | 8 | 0 | 39 | 23 | 2 | 14 | 18 | 0 |
| Tan-SH2 + FL SHP2 | 2 | 7 | 0 | 2 | 7 | 0 | 22 | 17 | 0 |
| SHP2 cat + FL SHP2 | 11 | 11 | 0 | 6 | 9 | 0 | 14 | 16 | 0 |

**Table 5.1 Analysis of fragment screening cascade showing the number of hits identified against the SHP2 catalytic, Tan-SH2 and full length SHP2 constructs by SPR, STD and waterLOGSY in the presence of HEPES and phosphate buffer conditions.** Fragment hits recorded for SPR and NMR-based screening of SHP2 catalytic domain (SHP2 cat), tandem SH2 domain (Tan-SH2) and full length SHP2 (FL SHP2) in HEPES and phosphate buffer conditions. Fragments confirmed as hits in both screening conditions by the same technique are also included. *denotes two cross-validated fragment hits against the SHP2 catalytic domain by STD and waterLOGSY in both HEPES and NaP buffer conditions (for SPR) and d-Tris and NaP buffer conditions (STD and waterLOGSY). Cross-validated hits for the SHP2 catalytic domain were then progressed to SPR dose-response experiments for further validation.

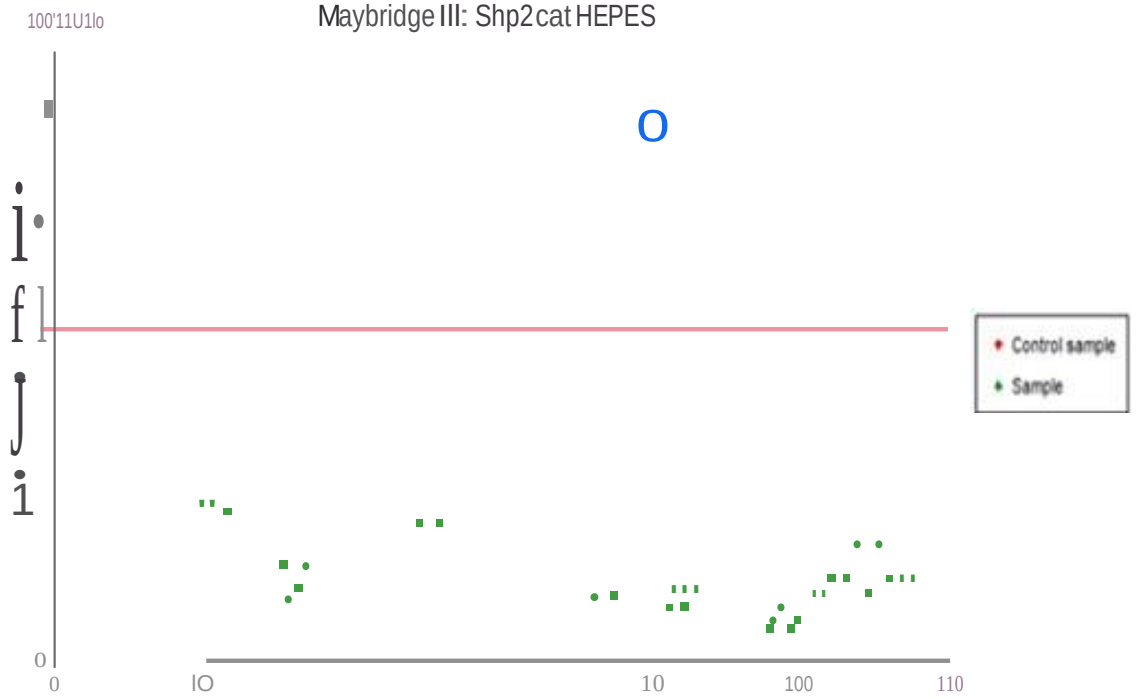
A


IIlyiHtdgt1: Shp2 cot HEPES
f·
j'
O
O
O
O
[;-]
It
IIIIII

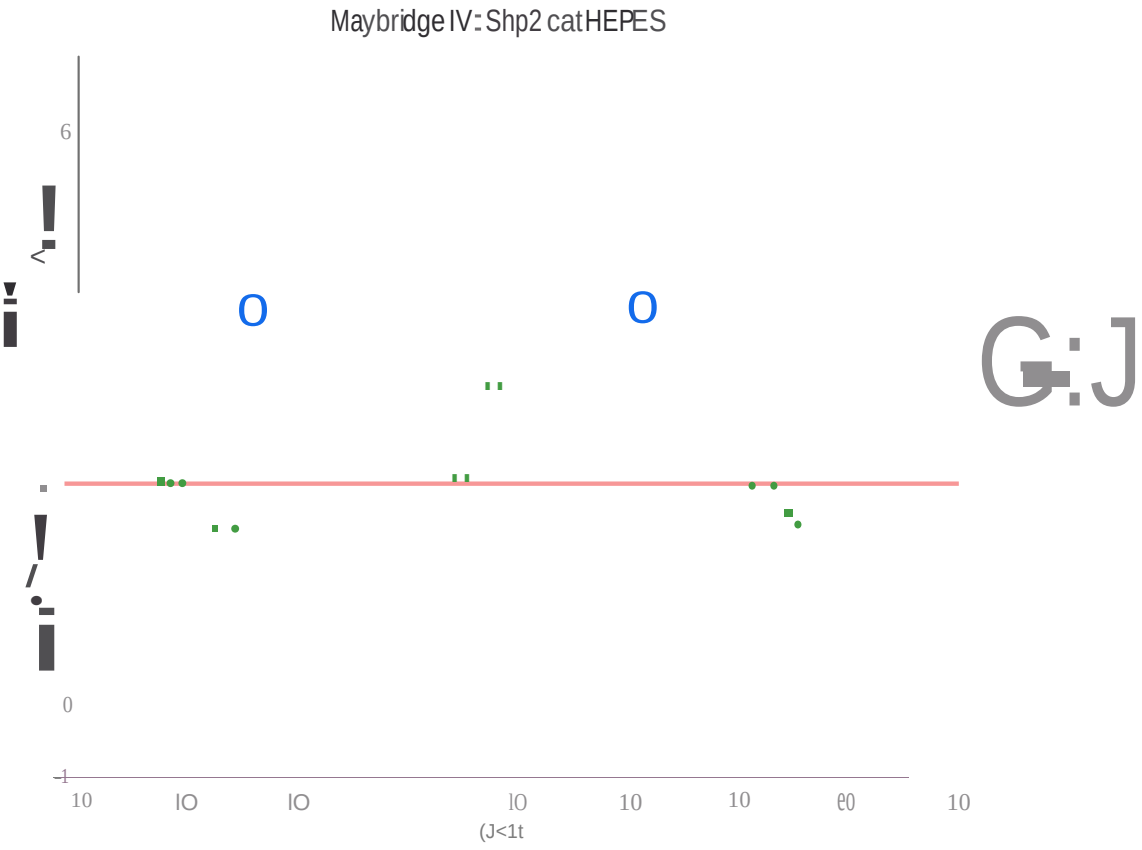
8


Maybridgt Il; Shp2 tit HEPES
L
1
10
10
C!<1t
Control sample
Sample

C


100'11U1lo
Maybridge III: Shp2 cat HEPES
Control sample
Sample
0
0
IO
10
100
110

D


Maybridge IV: Shp2 cat HEPES
6
0
-1
10
IO
IO
lO
10
10
e0
10
(J<1t

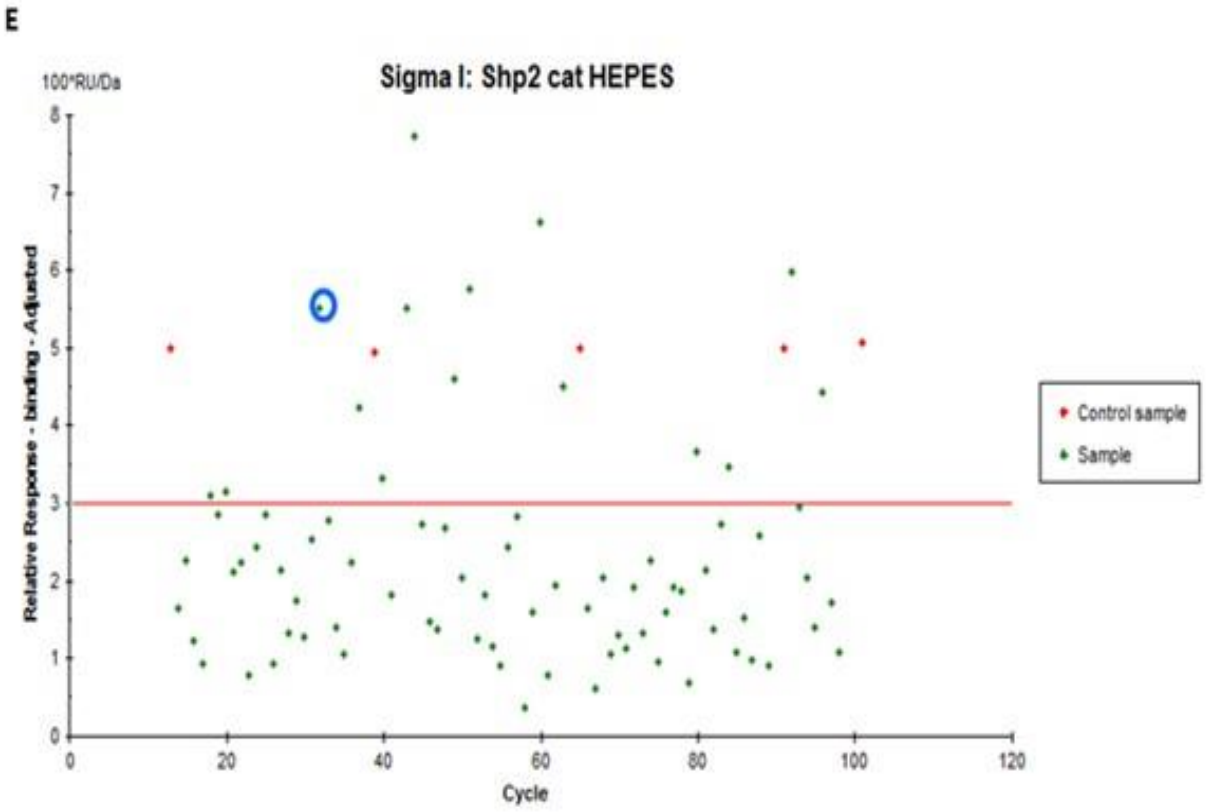



**Figure 5.7. SPR screening of a 433-compound fragment library against the SHP2 catalytic domain in HEPES buffer conditions**. Fragment hits were determined based on interrogation of individual sensorgram profiles giving responses at or above the 3 RU cut-off threshold. Fragments were screened as singletons at a single concentration of 200 uM in 5% DMSO. All 433 fragments were screened as singletons by dividing fragments into 5 separate screening runs. Figures A-E represent 5 individual screening runs for a single 433 fragment library screen. Labels correspond to chemical vendors from which the fragments were originally purchased. A - Maybridge I, B - Maybridge II, C - Maybridge III, D - Maybridge IV, E - Sigma I. Fragments demonstrating "table-top" sensorgram profiles were classified as hits. Fragments hits for each screen are circled in blue.

### 5.2.5 Fragment library screening by STD-NMR and waterLOGSY

As an orthogonal screening method both STD-NMR and waterLOGSY were used to counter screen the 433 compound fragment library in order to observe cross-validation of fragment hits highlighted in the primary SPR screen. These ligand-observed NMR techniques operate through directly monitoring the changes in ligand resonances in solution thus avoiding the need for protein immobilisation which could perturb native conformational dynamics and allosteric inhibitory mechanisms. For STD-NMR and waterLOGSY, 384 fragments from the original 433-compound library (49 fragments were excluded due to poor solubility) was screened in cocktails of 4 with each fragment present at a single concentration of 500 μM. The high fragment concentration (compared with the 200 μM concentration used in the previous SPR screens) was employed to ensure all ligand resonances were resolvable enabling accurate fragment hit determination. In comparison to the SPR screens, STD identified 27 initial fragment hits for the SHP2 catalytic domain in d-Tris buffer conditions giving a hit rate of 7%, compared to the same screen in phosphate which yielded 14 initial fragment hits and an overall hit rate of 3.6%. In contrast to the SPR screens, 9 fragment hits were confirmed by STD against the Tan-SH2 domain in d-Tris and 10 fragment hits in phosphate buffer conditions. waterLOGSY experiments identified a total of 24 fragment hits against the SHP2 catalytic domain in d-Tris buffer conditions giving a hit rate of 6.2%, compared with 14 fragment hits identified in phosphate buffer conditions giving a hit rate of 3.6%. For the Tan-SH2 domain,a total of 13 fragment hits were identified by waterLOGSY giving a hit rate of 3.3%, whereas screening of the Tan-SH2 domain in phosphate identified a total of 15 fragment hits against by waterLOGSY giving a fragment hit rate of 3.9%.

After completing SPR and NMR-based fragment screens against the SHP2 catalytic, Tan-SH2 and full length SHP2 domains, initial analysis of the SPR and NMR fragment screening data for the SHP2 catalytic domain revealed an unexpected and surprising lack of cross-validated hits, with only one fragment showing activity in both assay screens. In the interests of time and in line with project objectives, it was decided that further work would concentrate on providing further validation and characterization of SHP2 catalytic domain fragment hits from the SPR screen as chemical starting points for further optimization.

### 5.2.6 SPR dose-response analysis of fragment binding against SHP2 catalytic domain

To identify false-positive hits among the 11 fragment hits identified from initial SPR screening in HEPES buffer, these fragments were tested against the SHP2 catalytic domain in dose- response using concentration ranges from 6.4 µM-10 mM in 5-fold dilution steps. Of the 11 fragment hits identified by SPR screening, we identified the 2 active fragments 1A9 and 4B1 both of which displayed dose-response behaviour against the SPR catalytic domain typical of specific binding. The remaining 9 fragments were excluded from further analysis due to a combination of non-specific promiscuous binding, concentration-dependent aggregation, superstoichiometric binding (>3-fold ratio of ligand to protein), or weak binding which generated insufficient signal responses. After identifying fragment hits 1A9 and 4B1, both of these fragments were then repurchased and rescreened as singletons in order to independently confirm dose-response activity against the SHP2 catalytic domain and gauge the reproducibility of the dose-response results. In order to monitor the stability of the SHP2 catalytic domain and ensure dose-dependent responses of fragments were measured against active protein, dose-response analysis of NSC-87877

was assessed alongside the active fragments as a positive control compound. Indeed, NSC-87877 bound with the expected affinity and kinetics (previously reported in chapter 4) suggesting that SHP2 catalytic domain was active on the SPR sensor chip surface. Analysis of the dose-response curves reconfirmed that repurchased fragments 1A9 and 4B1 bound in a concentration-dependent fashion enabling the derivation of binding affinities for each fragment by fitting to a 1:1 ligand: protein stoichiometry. Fragment 1A9 was observed to bind SHP2 catalytic domain with a $K_D$ of 960 µM (and with an error in $K_D$ +/- 96 µM; Figure 5.8) and fragment 4B1 bound with a $K_D$ of 2 mM (with an error in $K_D$ of 0.2 mM; Figure 5.9).

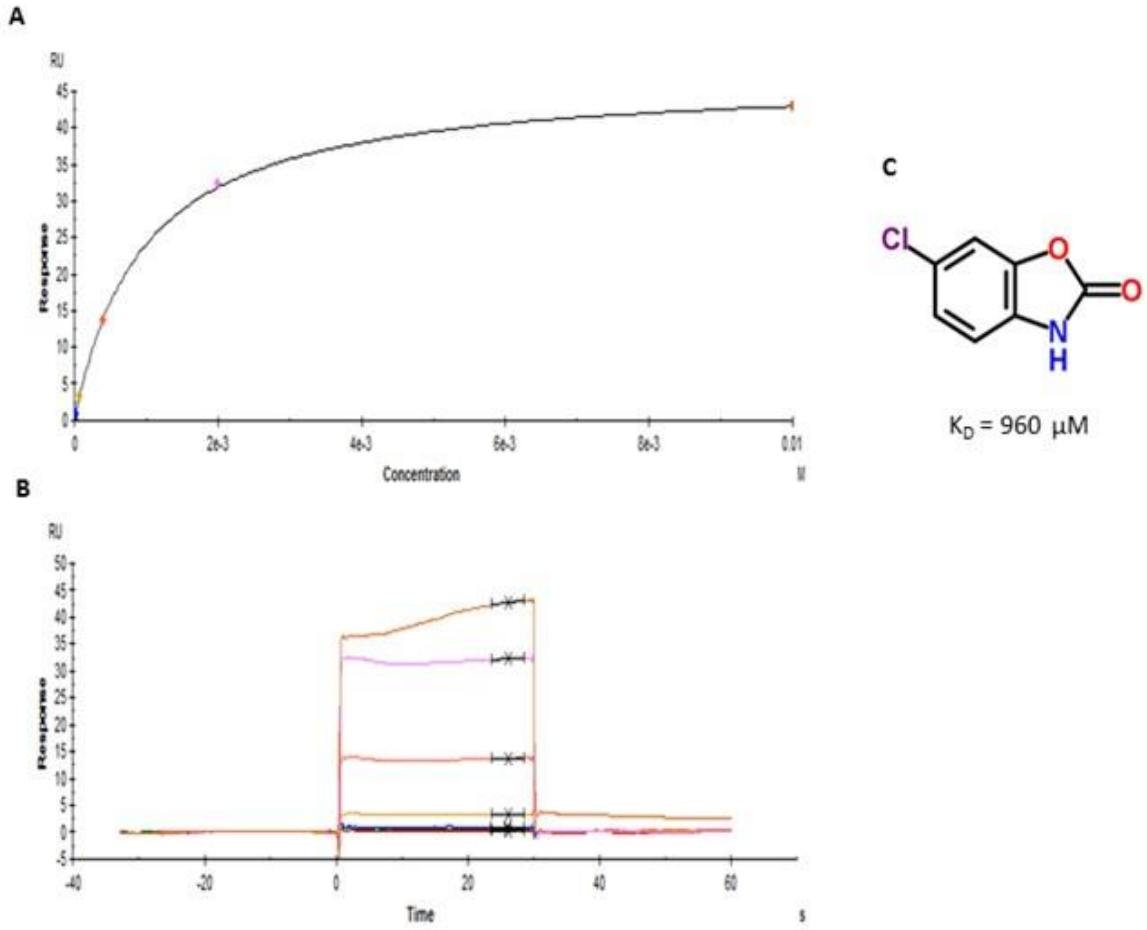



**Figure 5.8. SPR dose-response analysis for fragment 1A9 binding to SHP2 catalytic domain**. Five-fold serial dilutions of fragment 1A9 were tested across an eight point series in the concentration range of 0.64 µM - 10 mM and responses were fitted to the Langmuir model assuming a 1:1 protein: ligand binding stoichiometry. A $K_D$ for the binding interaction was determined as 960 µM.

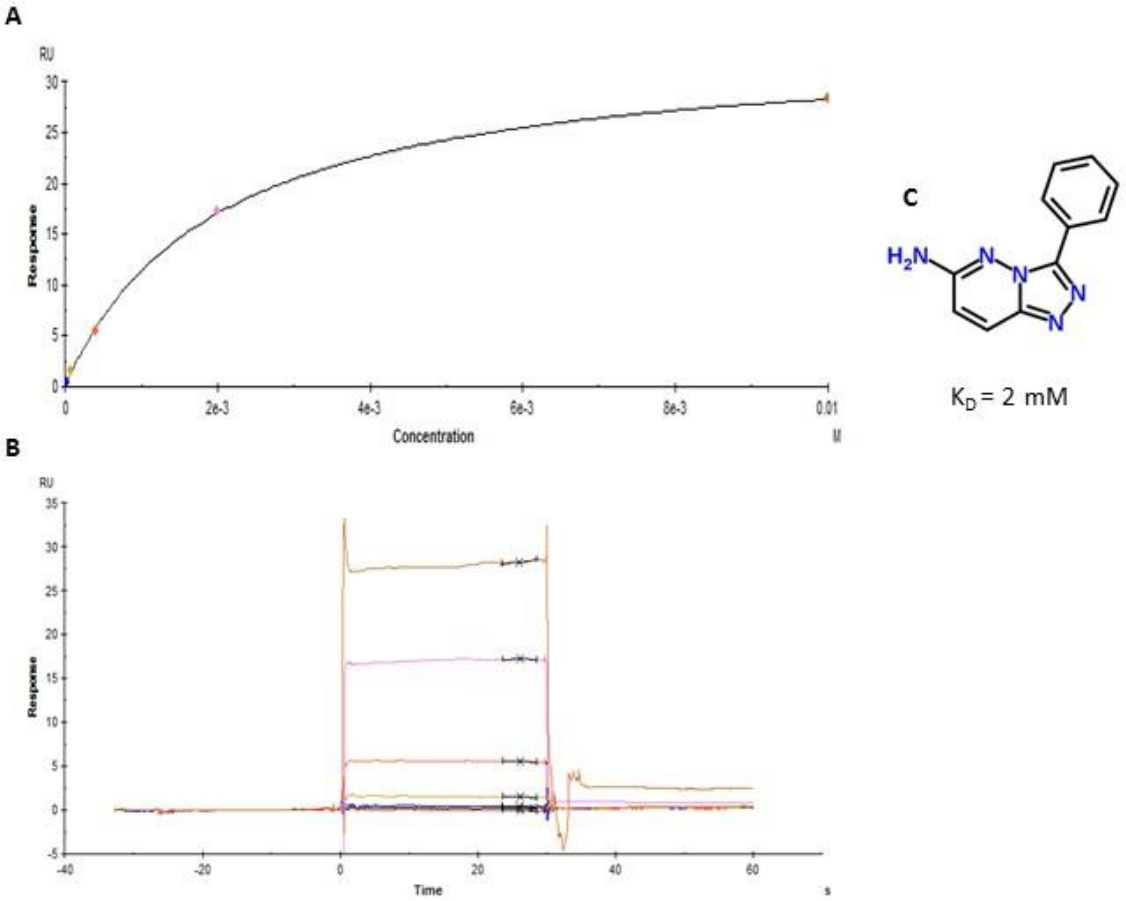



**Figure 5.9. SPR dose-response analysis for fragment 4B1 binding to SHP2 catalytic domain**. Five-fold serial dilutions of fragment 4B1 were tested across an eight point series in the concentration range of 0.64 μM - 10 mM and responses were fitted to a Langmuir model assuming a 1:1 protein: ligand binding stoichiometry. A $K_D$ for the binding interaction was determined as 2 mM.

### 5.2.7 SPR dose-response analysis for fragment binding to full length SHP2

After confirming the activity of fragments 1A9 and 4B1 against the SHP2 catalytic domain, a similar dose-response analysis of these fragments was then performed against full length SHP2. Under basal conditions, full length SHP2 exists in an auto-inhibited conformation whereby the N-SH2 domain conformationally occludes the active site [80]. Given the purposeful engineering of the Avi-tag™ at the C-terminal tail of the full length SHP2 SPR construct, it was assumed that the native auto-inhibited state would be the predominant conformational state of the protein immobilized on the SPR sensor chip surface. Thus, the next aim was to investigate whether fragments 1A9 and 4B1 both reproduced the same binding profile when binding to conformationally occluded catalytic domain compared with the previously binding profiles recorded against the open catalytic domain. This would yield insights into the binding location of each fragment and permit further exploration of novel fragment binding modes at unique pockets positioned at the interface between the N-SH2 domain and the catalytic domain of the full length SHP2 protein. Testing fragments 1A9 and 4B1 in dose-response against full length SHP2 (Figures 5.10 and 5.11) did not recapitulate binding profiles comparable with fragments binding to the open catalytic domain, suggesting that fragment binding is disrupted through N-SH2 domain-mediated occlusion of the phosphatase active site. Furthermore, the most likely explanation underlying the sensorgram profiles obtained for fragment 1A9 and 4B1 binding to full length SHP2 is that full length SHP2 occupies a mixed population of open and closed states on the SPR sensor chip. This suggests that the binding profiles obtained may represent an average of the fragment binding to both open and conformationally occluded full length SHP2.

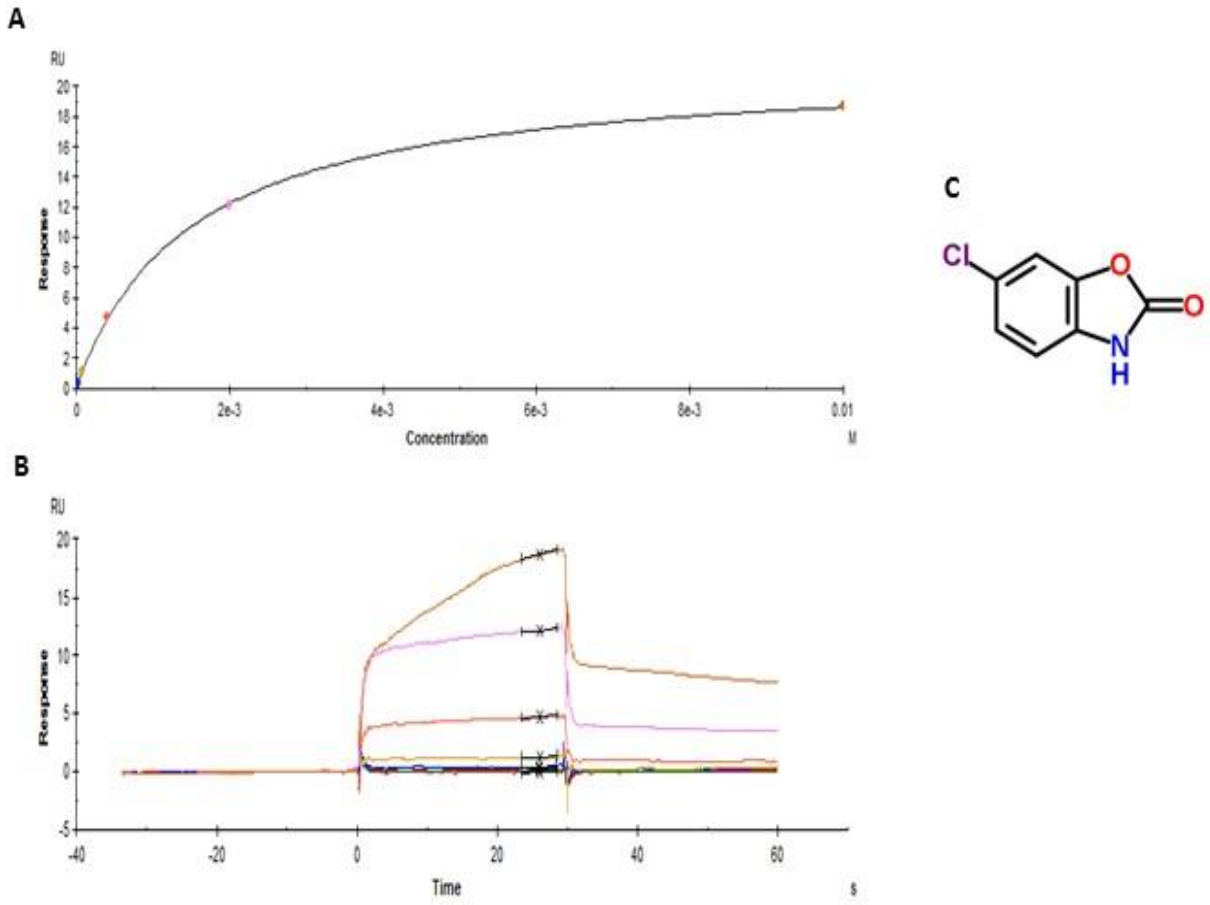



**Figure 5.10. SPR dose-response analysis for fragment 1A9 binding to full length SHP2.** Five-fold serial dilutions of fragment 1A9 were tested across an eight point series in the concentration range of 0.64 μM - 10 mM and responses were fitted to a Langmuir model assuming a 1:1 protein: ligand binding stoichiometry. Although an apparent $K_D$ was calculated as 1.2 mM, the response vs time sensorgram does not recapitulate the fragment binding profiles recorded previously for the open SHP2 domain. The fragment-binding profiles obtained suggest that the N-SH2 domain of SHP2 conformationally occludes the active site to perturb optimal fragment binding. Furthermore, the binding profile suggest that there is likely to be a mixed population of open and closed full length SHP2 immobilised on the chip, with the resulting sensorgram profile representing an average of the fragment binding response to both conformationally open and closed forms of full length SHP2.,

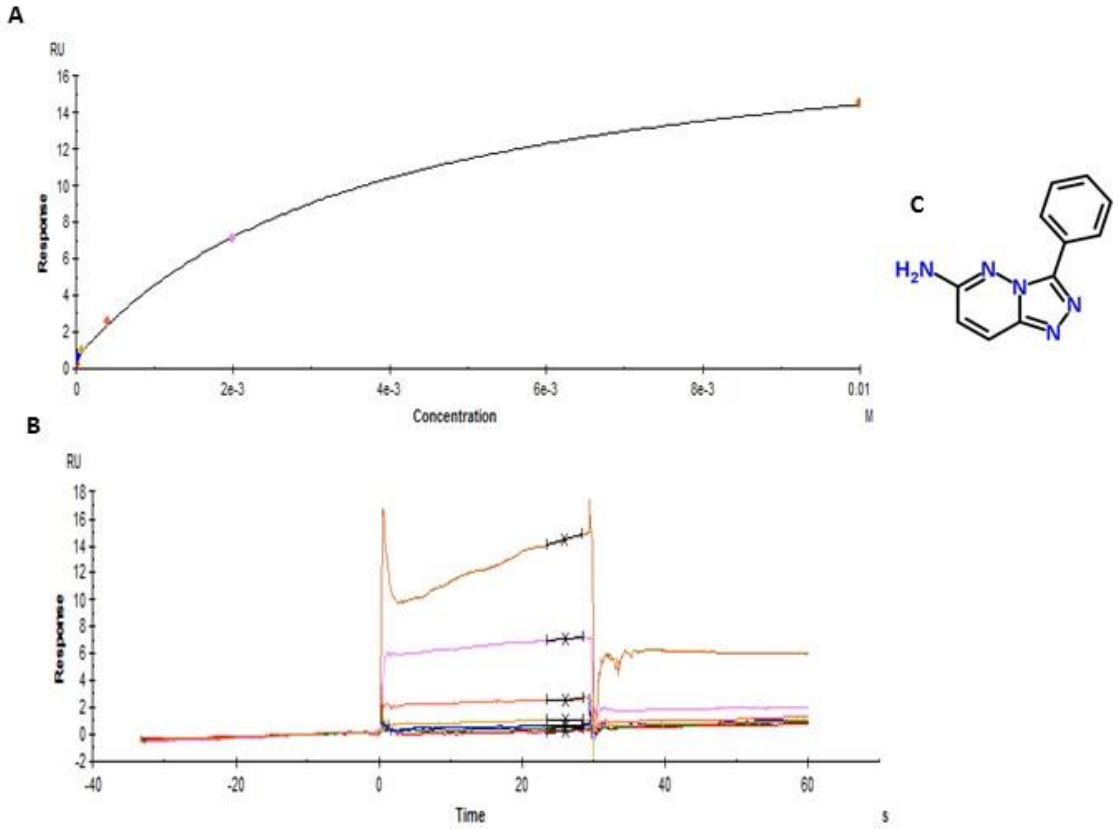



**Figure 5.11. SPR dose-response analysis for fragment 4B1 binding to full length SHP2.** Five-fold serial dilutions of fragment 4B1 were tested across an eight point series in the concentration range of 0.64 μM - 10 mM and responses were fitted to a Langmuir model assuming a 1:1 protein: ligand binding stoichiometry. Although an apparent $K_D$ was calculated as 2.5 mM, the binding response was insufficient to reach a steady state meaning an accurate $K_D$ value could not be derived. A similar profile for the binding response compared to that recorded for fragment 1A9 is consistent with the idea that the obtained sensorgram binding profiles represent an average of fragment binding to a mixed population of both conformationally open and occluded full length SHP2.

### 5.2.8 Cross validation of fragment binding by STD-NMR and waterLOGSY

To provide corroborative evidence for the dose-dependent binding of fragments 1A9 and 4B1 to the SHP2 catalytic domain by SPR and provide further validation of these fragments as chemical starting points for further development of inhibitors targeting the SHP2 catalytic domain, the repurchased fragments were retested in isolation by STD-NMR and waterLOGSY experiments. Although both the STD-NMR and waterLOGSY experiments gave particularly weak ligand signals during the initial fragment library screen, for the repurchased singletons of these fragments, the emergence of clear positive STD signals in alignment with chemical shifts recorded in $^{1}$H NMR spectra were observed for both fragments 1A9 (Figure 5.12A) and fragment 4B1 (Figure 5.13A) in the subtracted STD spectra, indicative of a binding interaction between fragments 1A9 and 4B1 and the SHP2 catalytic domain. Similar observations were also recorded in waterLOGSY spectra, with an inversion of the waterLOGSY signal upon addition of both fragment 1A9 (Figure 5.12B) and fragment 4B1 (Figure 5.13B) consistent with an interaction of these fragments with the SHP2 catalytic domain.

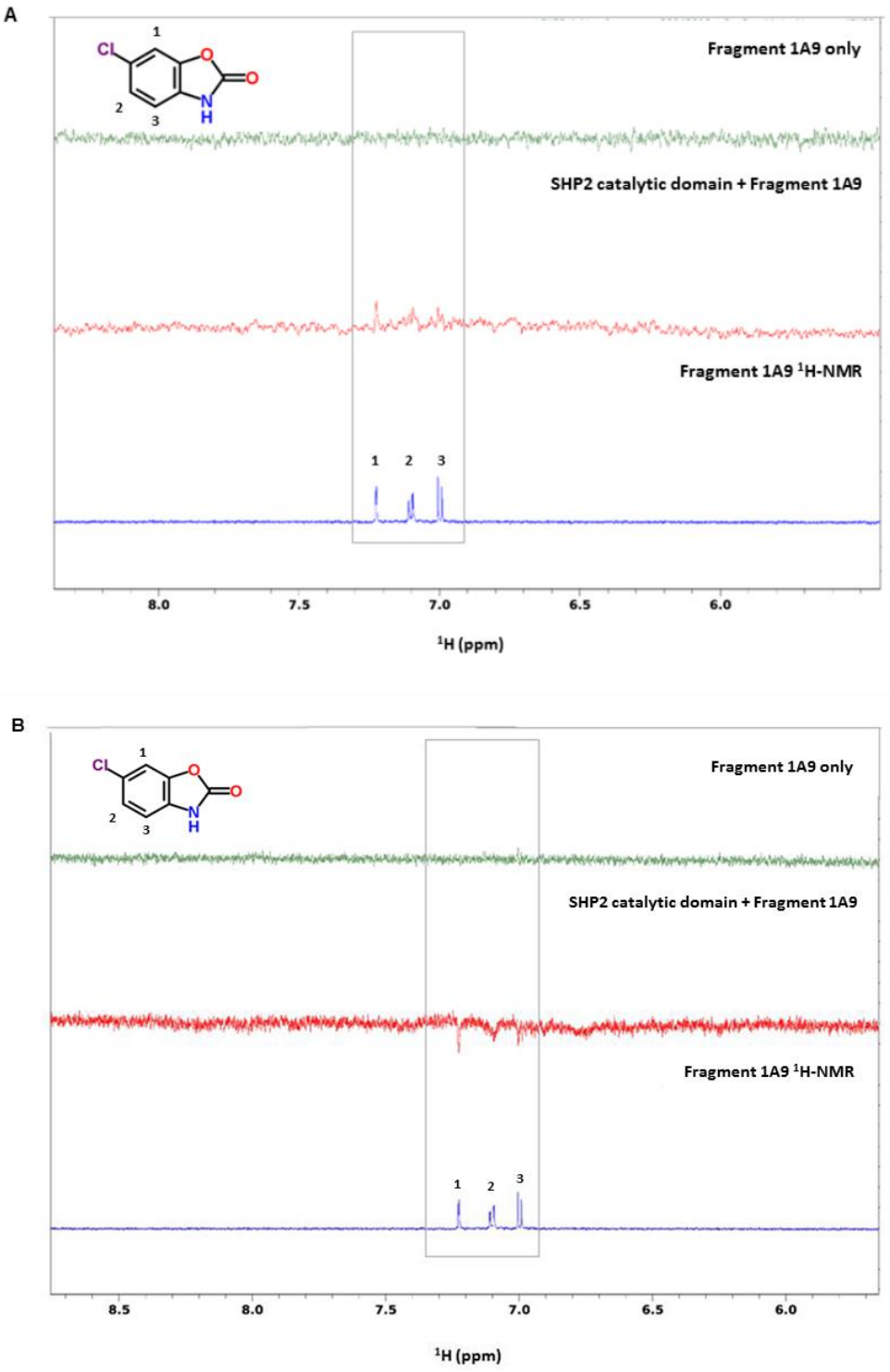



**Figure 5.12 Interaction of fragment 1A9 with the SHP2 catalytic domain by STD-NMR and waterLOGSY**. A. STD spectra of fragment 1A9 interaction with SHP2 catalytic domain, as shown by STD signals in substracted spectra (red). B. waterLOGSY spectra of fragment 1A9 interaction with SHP2 catalytic domain, as shown by inversion of waterLOGSY signal (red)

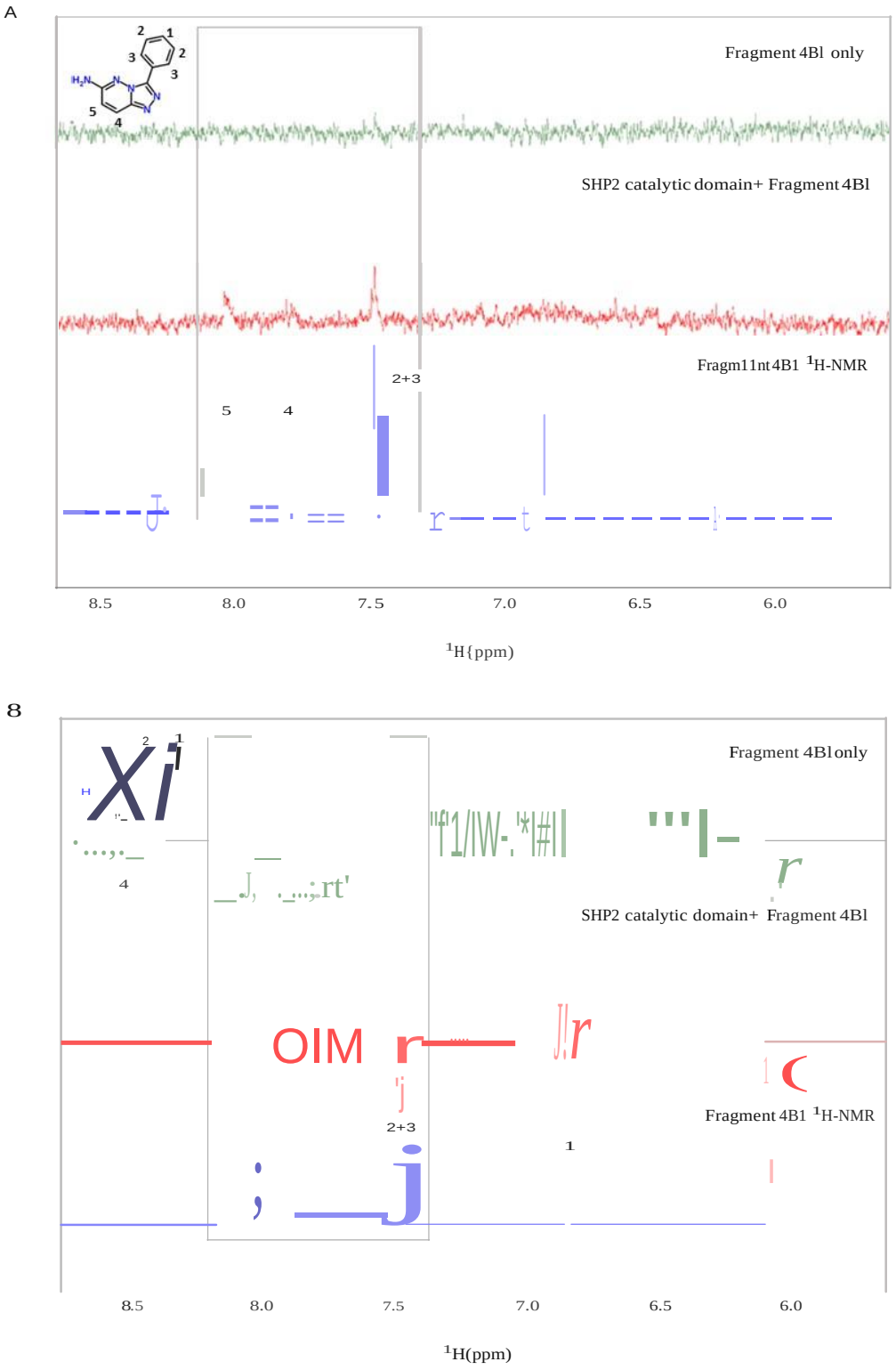



Figure 5.13 Interaction of fragment 4Bl with the SHP2 catalytic domain by STD-NMR and waterLOGSY. A STD spectra of fragment 4B1 interaction with SHP2 catalytic domain, as shown by STD signals in subtracted spectra (red). B. waterLOGSY spectra of fragment 1A9 interaction with SHP2 catalytic domain, as shown by inversion of waterLOGSY signals (red).

## 5.3 Discussion

Screening a 433-compound fragment library has enabled the successful identification of two chemically diverse fragment scaffolds which show orthogonal activity against the SHP2 catalytic domain by SPR, STD-NMR and waterLOGSY in HEPES buffer conditions. Alongside the SHP2 catalytic domain, the overall strategy also included screening of the Tan-SH2 domains and wild-type full length SHP2 protein in an attempt to identify fragments exploiting novel allosteric binding sites. Unfortunately, however, due to experimental time constraints further validation of fragment hits against the Tan-SH2 domain and full length SHP2 protein was not possible. Thus, efforts were focused on validation and characterization of hit matter solely for the SHP2 catalytic domain.

The primary screening assays for SPR and NMR-based (STD-NMR and waterLOGSY) fragment screening of the SHP2 catalytic domain were conducted in both HEPES and phosphate buffer conditions. This was decided based on preliminary findings reported in Chapter 3 which suggested that the presence of phosphate was necessary for stabilization of the overall structure of SHP2 enabling greater resolution of backbone amide signals in solution (as demonstrated by $^{15}$N-$^{1}$H TROSY HSQC studies on $^{15}$N-labelled SHP2 catalytic domain). Although the presence of phosphate may prohibit the binding of active-site directed fragments, it was hypothesized that the additional stabilization provided by phosphate enriched-screening conditions may expose novel and previously unexploited pockets for fragment binding allowing characterization and exploitation of novel binding sites for SHP2 inhibitor development. However, when comparing the SPR and NMR screening results in HEPES which exhibited 2.5% and 7% fragment hit rates respectively against the SHP2 catalytic domain, the identical screen of SHP2 catalytic domain conducted in the presence of phosphate buffer exhibited 1.3% and 3.2% fragment hit rates

for the SPR and NMR screens, suggesting that the presence of phosphate in the screening buffer has no significant effect on fragment hit rates for the SHP2 catalytic domain or the type of fragments (in terms of particular chemical scaffolds) that bind for the SHP2 catalytic domain. Since screening phosphatases by fragment-based methods remains a highly underexplored discipline, it is not yet clear whether this observation is specific for screening studies involving SHP2 alone or whether such an approach can be applied more generically to guide future fragment-based screening efforts against other phosphatase targets

The 2.5% fragment hit rate (11 fragment hits from a total of 433 fragments screened) obtained from SPR screening of the SHP2 catalytic domain at a single fragment concentration of 200 μM recorded in HEPES buffer is low compared to hit rates generally observed for fragment-based screening efforts, which typically yield hit rates between 3-10% [222]. However, fragment hit rates are dependent on a variety of factors, the most important of which include the size and diversity of the fragment library, and have also been suggested as a useful metric for the assessment of the druggable potential of a particular target [223]. Indeed, despite the relatively underexplored area of phosphatase drug discovery, the notion of fragment hit rates correlating with the overall druggability of a particular target is supported by recently reported fragment screening efforts against phosphatases. For example, screening an in-house library of 1500-chemically diverse fragments against the phosphatase CDC25B by $^{15}N$-$^{1}H$ NMR spectroscopy identified a total of 1 fragment hit, giving a hit rate of 0.05% [163]. Compared with other fragment libraries, the 433 compound fragment library used in this work is significantly smaller, with many commercially and industrially developed libraries containing in the range of 1000-3000 fragments [164] . Thus, the relatively small size of the fragment screening library

combined with the renounced and poor druggability of phosphatases are both key factors underlying the relatively low hit rate of 2.5%.

From the SPR screening assay, a total of 11 fragment hits were recorded from the screen conducted in HEPES buffer, which after dose-response analysis was reduced to 2 active fragments, 1A9 and 4B1. Interestingly, the parallel SPR screen conducted in phosphate buffer recorded a total of 6 fragment hits, of which 4 fragments overlapped with the 11 hits against the SHP2 catalytic domain recorded in HEPES. Of these 4 overlapping hits, the validated fragment 1A9 bound in both HEPES and phosphate screening conditions, whereas fragment 4B1 was identified only in the HEPES screen. Although further studies are necessary in order to validate the binding locations of these fragment hits, abrogation of fragment 4B1 binding in the presence of phosphate could be suggestive of a binding site located in SHP2 active site. In contrast, fragment 1A9 binding was independent of the buffer screening condition suggesting that a putative binding site for this fragment could be either the SHP2 active site or an alternative binding location distinct from the phosphatase active site. Multiple experimental methods can be utilized for the validation of fragment binding locations. These include, but are not exclusive to X-ray crystallography [224], 2D-NMR [225] and STD-competition experiments, all of which offer complementary but differing levels of information regarding fragment binding sites. Unfortunately, as documented in Chapter 3, the high phosphate concentrations required for optimal resolution of $^{15}N$-$^{1}H$ amide signals poses challenges for mapping inhibitor binding sites by 2D-NMR methods. For this reason, it was deemed that the weak binding affinities of fragments 1A9 and 4B1 (as determined by SPR) coupled with the high fragment concentrations required for binding site mapping would compromise fragment solubility and thus be insufficient to overcome the saturating levels of phosphate occluding the SHP2

active site. In contrast, although STD-fragment competition experiments were attempted towards the latter part of this study, time constraints meant that it was not possible to optimize the experimental parameters to a degree that enabled observations of displacement of fragment 1A9 and 4B1 binding in the presence of the NSC-87877 competitor ligand. Indeed, STD-fragment competition assays have been routinely employed with great success as part of kinase drug discovery drug initiatives, which have greatly benefited from the use of the potent active-site competitor ligand strausporine [226].

Despite preliminary attempts to identify the binding locations of these fragments, confirmation of fragment 1A9 and 4B1 binding was observed by dose-response analysis which revealed both fragments bound with affinities of 960 µM and 2 mM against the SHP2 catalytic domain, respectively. Indeed, the weak affinities of these fragments explain why it was not possible to observe these fragment hits when conducting the preliminary fragment library screens at a single concentration of 50 µM. Interestingly, despite identifying fragments 1A9 and 4B1 as initial hits in the SPR screen, these fragments were missed as hits in the orthogonal STD and waterLOGSY screens. Furthermore, when comparing the level of cross-validation between fragment hits recorded by SPR, STD-NMR and waterLOGSY, surprisingly only one fragment hit out of a total of 433 fragments was confirmed as a successful hit using all three methods. It was only through repurchasing of the 11 SPR fragment hits as singletons and the subsequent retesting of fragments 1A9 and 4B1 in dose-response experiments that provided confirmation of the orthogonal activity of these fragments by STD-NMR and waterLOGSY. Although the lack of cross-validated hits against the SHP2 catalytic domain from the SPR and NMR-based screens was initially surprising, similar observations have also been reported using SPR and NMR-based fragment screening methodologies for alternative protein target classes.

For example, fragment screening campaigns conducted against the HIV-1 integrase core domain have reported similar observations revealing an unexpected lack of hit cross-validation when utilizing SPR and STD-NMR as primary screening tools [227]. In that study, Wielens et al. conducted two independent fragment screens using 500 compound fragment libraries selected from the Maybridge collection, both of which showed only 10% variation in chemical context and shared 455 identical fragments. The two screening cascades utilized different screening methodologies for the initial detection of fragment hits; one fragment library was screened using SPR, and the other library screened using STD-NMR. Interestingly, despite screening virtually identical fragment libraries, the analysis revealed very little cross-correlation between fragment hits using these methods. The authors posited the differences in experimental conditions used in each screening cascade as a possible explanation underlying this observation, citing differences in pH (pH 7.5 in SPR screen versus pH 8.5 used in the STD-NMR screen), DMSO concentration (5% (v/v) DMSO used in SPR screen versus 1.5% (v/v) DMSO used in STD-NMR screen), salt concentration (150 mM NaCl in SPR screen versus 50 mM in STD-NMR screen) and fragment concentration (0.5 mM in SPR screen vs 1 mM in STD-NMR screen) between the two screens as possible factors contributing to the lack of hit cross-validation. However, rescreening the initial hits from the primary screens at the same pH revealed no significant differences in hit overlap between the two screens. This led the authors to suggest that the lack of observable hit overlap between the two screens was primarily a result of the difference in hit selection criteria rather than differences in screening conditions. For example, in the STD-NMR screen the authors selected even fragments giving weak STD signals in order to enrich the pool of novel fragment chemotypes for further evolution and medicinal chemistry optimisation. However, in the SPR screen, the

authors incorporated a cut-off value for hit determination which prioritized the selection of fragment hits at or above the calculated cut-off value [227].

The observations reported by Wielens et al. form an explanatory framework which can be used to interpret the results from fragment-based screening of SHP2 described in this chapter. In a similar manner, an underlying explanation for the lack of cross-validated fragments hits observed during SHP2 screening is likely a result of the different sensitivities of NMR and SPR-based screening methods. Furthermore, screening at different fragment concentrations by SPR (0.2 mM) and STD-NMR and waterLOGSY (0.5 mM) encourages selection of two different fragment populations where higher affinity fragments are preferentially selected as hits by SPR (and are often overlooked as hits in the STD-NMR and waterLOGSY screens due to weaker signals), and lower affinity fragments that fall under the SPR cut-off are detected by the STD-NMR and waterLOGSY screens.

## 5.4 Conclusion

Utilizing SPR and NMR-based fragment screening methods (STD-NMR and waterLOGSY) has enabled the identification of novel and chemically distinct fragments hit against the SHP2 catalytic domain which now serve as starting points for further validation and characterization. More specifically, fragment screening assays combined with further validation of fragment hits enabled deconvolution of hit matter from 433 fragments to confirm fragments 1A9 and 4B1 as dose-response active fragments against the SHP2 catalytic domain, which bind with affinities of 960 μM and 2 mM respectively. Similar dose-response analysis of these fragments against conformationally occluded full length SHP2 was unable to reproduce the sensorgram binding curves comparable with similar dose-response analysis against the open catalytic domain. The explanation for this

phenomenon is likely to be a result of a mixed population of open and closed full length SHP2 immobilised on the SPR chip, suggesting that the fragment binding profiles recorded represent an average of fragment binding to two different populations of conformationally open and occluded forms of full length SHP2.

# CHAPTER 6: DETERMINING A LIGAND SOAKABLE CRYSTAL FORM OF THE SHP2 CATALYTIC DOMAIN

## 6.1 Experimental aims

Fragment-based library screening of the SHP2 catalytic domain by SPR and NMR-based methods enabled the successful identification of chemically distinct fragments 1A9 and 4B1 with binding affinities of 960 µM and 2 mM for the SHP2 catalytic domain, respectively (Chapter 5). These validated fragment hits now serve as novel chemical starting points for optimisation and development of more potent SHP2 inhibitors. To aid the iterative evolution of these fragments into more potent and selective inhibitors, our aim was to determine the x-ray crystal structures of these fragments in complex with the SHP2 catalytic domain to provide experimental validation of fragment binding modes and utilise this structural information to guide the medicinal chemistry-driven expansion and progression of these fragments into more potent lead inhibitors. In order to facilitate this structure-guided approach, the over-arching aim of this was to optimise and develop a sufficiently robust and soakable crystal form of the SHP2 catalytic domain that was amenable for fragment-soaking and structure determination.

## 6.2 Results

The progression and evolution of validated fragment hits to facilitate the development of drug-like inhibitors is greatly accelerated by x-ray crystallographic structure determination of fragments bound to the corresponding protein target. Such high resolution structural information provides experimental validation of fragment binding modes, can illuminate novel and previously undisclosed druggable pockets and enables structure-guided elaboration of fragments through an iterative cycle of medicinal chemistry optimisation to evolve weak affinity fragment starting points into more potent lead inhibitors. In an attempt to optimise fragments 1A9 and 4B1 into more potent drug-like inhibitors of the SHP2 catalytic domain, it was necessary to devise a robust and soakable crystal system that would allow experimental validation of fragment binding modes as a precursor to medicinal chemistry optimisation. Due to the high concentration of phosphate required, which would inevitably result in saturation of the SHP2 active site, the use of $^{15}N$-$^{1}H$ HSQC NMR and observation of chemical shift perturbations was deemed an intractable approach for characterising fragment binding modes. The less physiologically relevant but higher resolution method of x-ray crystallography was therefore pursued in an attempt to crystallise an open and unoccluded form of the SHP2 catalytic domain that was amenable to fragment-soaking and permitted determination of fragment-bound crystal structures to resolutions sufficient for driving medicinal chemistry optimisation.

### 6.2.1 Expression and purification of SHP2 catalytic domain

Purification of recombinant human SHP2 catalytic domain (residues 247-534) was facilitated by a two-step method of $Ni^{2+}$-NTA and size-exclusion chromatography (as described in Materials and methods) which yielded sufficient quantities of stable, homogenous and TEV-cleaved protein for structural analysis at a final concentration of 41 mg/ml (Figure 6.1).

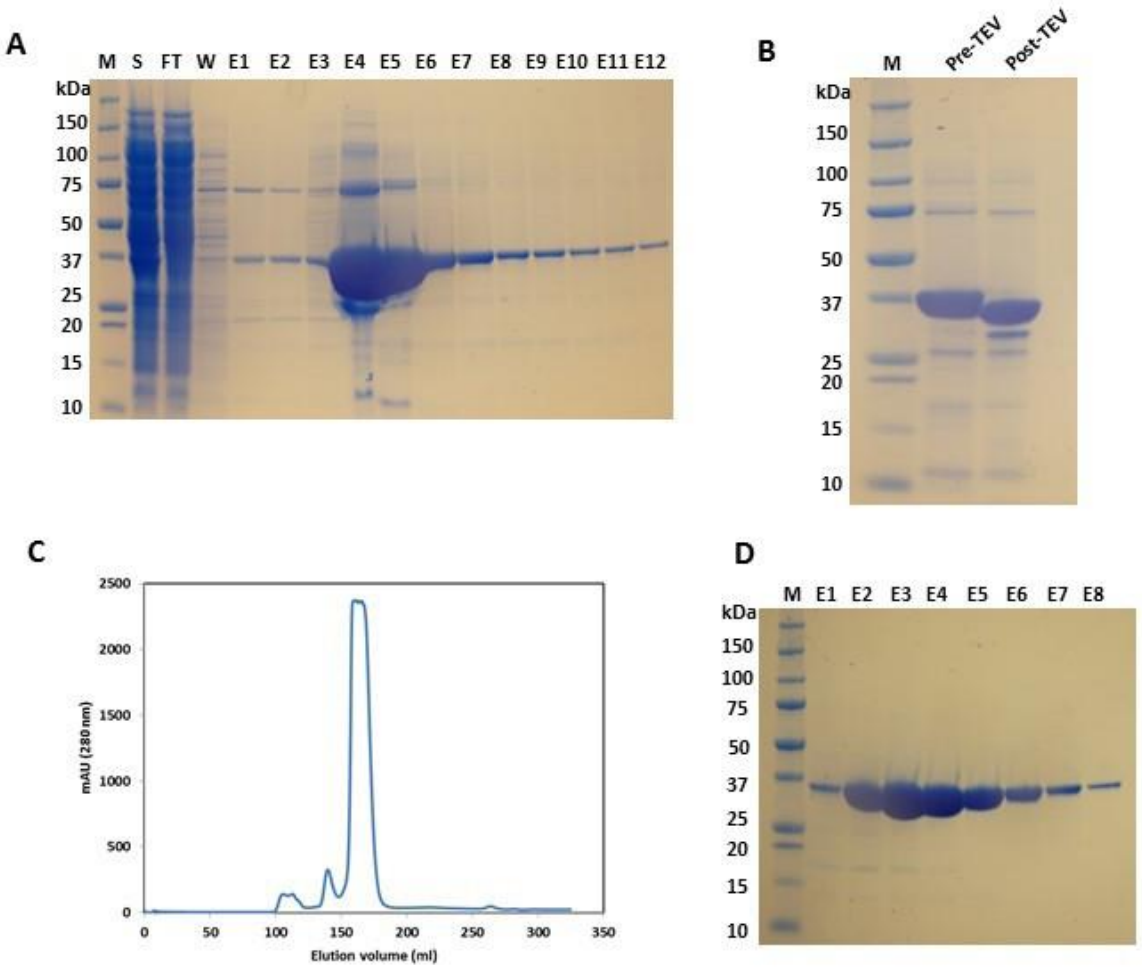



**Figure 6.1. Expression and purification of unlabelled SHP2 catalytic domain for crystallisation**. A) Purification of unlabelled SHP2 catalytic domain by $Ni^{2+}$-NTA affinity chromatography from 6L BL21 (DE3) cells. From left to right: Molecular markers (M), supernatant (S), flow-through (FT), Wash (W), Elution fractions 1-12 (E1-E12). B) Cleavage of Hisx6-tag pre- and post-incubation with TEV protease. From left to right: Molecular markers (M), pooled protein fractions prior to incubation with TEV protease (pre-TEV), pooled protein fractions after overnight dialysis and incubation with TEV-protease (post-TEV). C) Size-exclusion chromatography of His-cleaved SHP2 catalytic domain. SHP2 catalytic domain elutes as a single peak indicating the presence of a stable monomeric species. D) Purified fractions of His-cleaved SHP2 catalytic domain by size-exclusion chromatography.

Rather fortuitously, the expression construct of the SHP2 catalytic domain used for this investigation (donated as a kind gift by the SGC, University of Oxford) had been crystallised previously to 1.8 Å resolution and the co-ordinates and crystallisation conditions deposited in the protein databank under PDB code 3B7O [38]. Unfortunately, despite employing an extensive crystal screening strategy, preliminary attempts to identify alternative crystallisation conditions for the SHP2 catalytic domain were unsuccessful. Owing to the lack of success in the search for alternative crystallisation conditions, it was decided that the next strategy would involve exploiting the known crystallisation conditions used for structure determination of SHP2 catalytic domain (PDB code: 3B7O) and using these crystallisation conditions as a starting point for further crystal optimisation. In parallel to this objective, the secondary aim was to explore the tractability of this crystal system by assessing whether crystals generated from these conditions were sufficiently robust (tolerant to DMSO concentrations typically not exceeding 10%) and soakable (crystallise in an open form with an accessible binding site). Both crystal robustness and soakability (amongst other criteria to be discussed throughout this chapter) are the key criteria when assessing crystal suitability for fragment soaking [224].

### 6.2.2 Thermal shift assays to determine optimal buffer stability

Prior to setting up crystal trials we considered it prudent to assess the thermal stability of the SHP2 catalytic domain across a range of buffer conditions in order to identify the most stable buffer condition giving maximal protein stability. This we reasoned would yield a more homogenous population of stable protein in solution encouraging more efficient protein packing during crystallisation and leading to a more ordered crystal lattice. In order to assess thermal stability we utilised the Thermofluor$^{TM}$ system to assay the stability of

SHP2 catalytic domain across a range of different buffer conditions. The Thermofluor™ assay relies primarily on the changes in fluorescent properties of the hydrophobic dye SYPRO Orange which recognises and binds to hydrophobic patches of protein structure. Since hydrophobic regions of protein structure are typically localised to the core of most proteins (which is shielded from solvent exposure), an increase in fluorescence signal is often observed upon gradual increases in temperature. This drives protein unfolding ultimately leading to the exposure of hydrophobic patches and an increase in the propensity of dye binding [228]. In this way, buffer conditions giving fluorescent signals at the highest melting temperature correlate with optimum protein stability. The thermal stability assay therefore involved assessment of protein stability by recording thermal melting curves of protein unfolding for the SHP2 catalytic domain over 96 different buffer conditions. The experimental method for this assay is described in Materials and Methods. Analysis of the 96 thermal melt curves (data not shown) identified 50 mM HEPES pH 7.5 as the buffer condition producing the greatest fluorescence signals at higher temperature indicating that this condition was most conducive to optimum thermal stability. Therefore, prior to setting up crystallisation trials of SHP2 catalytic domain, SHP2 catalytic domain was dialysed into buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl and 10 mM DTT.

### 6.2.3 Overview of X-ray crystallography

X-ray crystallography is a widely used technique for determining the atomic resolution structures of protein molecules at the molecular level. The insights obtained from X-ray structure analysis can be utilised for the assignment of protein function, to aid understanding of substrate and inhibitor interactions with target proteins to unravel enzyme

mechanisms and accelerate drug development, and can also be used to probe the functional consequences of protein catalysis in different biological systems. Although X-ray crystallography has several caveats, namely that crystal structures exhibit static representations of protein structure rendering them unable to account for the dynamic flexibilities of protein molecules essential for driving catalysis, the resulting structures are often generated to an unprecedented level of accuracy. The process of protein structure determination by X-ray crystallography is a multi-step process, all of which will be summarised in forthcoming sections.

#### 6.2.3.1 Protein crystallisation theory

The primary and most fundamental requirement of protein structure determination by X-ray crystallography is the generation of highly-ordered, diffraction-quality crystals. This process typically involves an iterative cycle of optimisation and refinement and relies on the identification of crystallisation conditions yielding crystals of appropriate size and morphology for X-ray diffraction. In order to drive protein crystallisation, the overall aim is to create conditions which reduce protein interactions with the solvent whilst simultaneously maximising protein-protein interactions. To encourage protein-protein interaction leading to crystal formation, the protein solution needs to reach a supersaturated state which increases the probability of protein self-association. The overall process of protein crystallisation can be broadly separated into three distinct but related stages. These stages include crystal nucleation, growth and termination, and the sequential order of these processes underlies the growth of all crystals. The first stage of crystal nucleation refers to the early stages of crystal growth and often results in the formation of small microcrystal aggregates or “nuclei”, each of which can act as a seeding site to

encourage new crystal growth. The perpetual growth of the crystal is then driven by the recruitment of multiple protein molecules to sites of nucleation, which over time may result in larger and morphologically distinct crystals suitable for X-ray diffraction analysis. The final stage of the crystallisation process involves termination, where crystal growth is arrested often due to the limiting availability of protein molecules, growth defects, or the presence of contaminants which impede further crystallisation.

In order to achieve the super-saturated protein states required for crystallisation, conditions must be appropriately configured in order to reduce protein solubility. The protein solubility is influenced by a multitude of different factors which include pH, ionic strength and the presence of precipitants. In particular, pH and the protein isoelectric point (pI) are key parameters for determining the solubility of all proteins, with pH values closest to the isoelectric point (pI) representing conditions where protein solubility will be at its lowest. As such, for any given protein, the optimal crystallisation conditions will therefore be a combination of pH, ionic strength and precipitant concentration sufficient to compromise protein solubility and minimise protein-solvent interactions while simultaneously driving the unfavourable self-association of proteins to yield a diffraction-quality crystal.

#### 6.2.3.2 Protein crystallisation by vapour diffusion

The generation of protein crystals using the vapour diffusion method is the most commonly employed technique in protein crystallisation, and typically involves two primarily different approaches known as hanging drop and sitting drop methods. The use of one or both of these approaches in combination allows multiple different crystallisation conditions to be screened in parallel, with the ultimate aim of identifying initial screening conditions as a starting point for further optimisation of crystal growth. Methodologically,

both hanging drop and sitting drop approaches involve mixing equal volumes of protein solution with crystallisation buffer (usually precipitant plus additives) and then leaving this solution to incubate at 5$^{o}$C or 25$^{o}$C for crystallisation to occur. For the hanging drop method, crystallisation is usually conducted by mixing protein and crystallisation solutions onto the surface of a cover slip and then inverting the cover slip over a reservoir solution containing the corresponding crystallisation screening condition. For the sitting drop method, individual wells or sitting bridges are used to suspend the crystallisation drop above the reservoir solution. Despite differences in methodology, the thermodynamic principles underlying protein crystallisation by hanging and siting drop vapour diffusion are fundamentally similar. The underlying catalyst for driving crystal growth originates from the difference in precipitant concentration between the crystallisation drop and the reservoir solution. More specifically, the lower precipitant concentration in the crystal drop compared with the precipitant concentration of the reservoir solution causes the diffusion of water from the drop by vapour diffusion establishing an equilibrium between the drop and the solution. The diffusion of water from the crystallisation drop leads to a gradual increase in protein concentration which is eventually sufficient to reach the supersaturated state required for optimal crystallisation. After reaching a supersaturated state, the number of protein-protein interactions vastly supersede the number of protein-solvent interactions leading to the unfavourable self-association of protein molecules which acts as the trigger for early stage nucleation and crystal growth, or precipitation. The precipitation of protein in the crystallisation drop is primarily driven by the non-specific interaction between protein molecules which is insufficient to support crystal growth.

The preliminary stages of initial crystal screening are typically conducted using a 96-well plate format and benefit from the utilisation of a variety of commercially available sparse

matrix screens. Setting up individual screening conditions in this manner often involves nanolitre drop sizes set up in an automated fashion using a mosquito crystallisation robot, which allows screening to be performed on a larger scale and across multiple crystallisation conditions with relatively low protein consumption. The individual screening conditions are then left for crystallisation to occur which depending on the type of protein can typically take between 1-14 days. The buffer conditions yielding initial hits are then selected for further optimisation, with optimisation of initial hits typically involving adjustments to the protein concentration, buffer conditions, precipitant concentration or the buffer pH.

After optimising initial hit conditions to generate crystals of suitable size and morphology, crystals are then subjected to x-ray diffraction analysis to aid structure determination. However, prior to exposure to the x-ray beam crystals must be immersed in a suitable cryoprotectant solution in order to minimise damage induced by high energy x-ray radiation. The cryoprotectant, which acts by forming a protective barrier around the protein crystal, additionally serves to prevent ice crystal formation around the crystals which can lead to the unwanted and spurious inclusion of diffractions patterns and interference with data processing. The composition of the cryoprotectant solution usually consists of the reservoir solution from which the crystal was grown combined with additional PEG, glycerol or ethylene glycol, with the exact composition tailored to the precise recipe of the mother liquor. After development of an appropriate cryoprotection strategy, protein crystals are immersed in the cryoprotection solution before being flash frozen in liquid nitrogen ready for x-ray diffraction analysis.

#### 6.2.3.3 X-ray diffraction and data collection

To facilitate x-ray data collection, protein crystals previously flash frozen in liquid nitrogen must first be mounted in a nylon loop to ensure that the crystal position coincides with the direct path of the x-ray beam. Once mounted, the crystal is maintained at a constant temperature by a cryostream of nitrogen gas which also serves to preserve crystal integrity. The mounted crystal is then rotated periodically through a number of defined angles and bombarded with a beam of high intensity x-rays which interact at different orientations of the crystal in order to record a complete dataset. The high energy x-ray beam collides with the protein crystal and interacts with electrons in the protein molecules inducing oscillations which re-emit the x-rays as discrete beams. This generates a unique diffraction pattern on an x-ray detector which, following data processing and analysis, facilitates protein structure determination.

#### 6.2.3.4 X-ray data processing

The first objective of processing x-ray diffraction data as a precursor to structure determination involves defining the unit cell parameters and space group characteristics of the protein crystal. The unit cell is fundamentally defined as the simplest repeating structural unit of the crystal and therefore is an essential metric for measuring the degree of internal order and regularity. The unit cell parameters for any given crystal are divided into seven distinct categories, with each unit cell dimension defined by three cell lengths (a, b and c) and three corresponding angle assignments (α β, γ). Each of these unit cell categories exhibited by a crystal imposes a defined set of constraints on these length and angles. Therefore, in order to proceed with structure determination, the unit cell parameters of the crystal must be derived experimentally by calculating the distances between the

spots recorded in X-ray diffraction pattern. In conjunction with defining the unit cell parameters, the crystal space group must also be defined which provides a description of the precise arrangement of the protein molecules in the unit cell of the crystal.

Determination of the unit cell parameters from an x-ray diffraction pattern relies on exploiting the fundamental properties of x-rays exhibited by all wave forms. These properties include the amplitude, frequency and phase of the wave, all of which are essential prerequisites for protein structure determination. After completing an x-ray diffraction experiment, the frequency of the wave can be calculated from the wavelength of the x-ray beam, and the wave amplitude can be routinely established during data processing through measurement of the individual spot intensities and converting these to amplitude. However, the crucial phase information of the wave is not obtained during data collection and therefore has to be determined independently through further experimentation. This notion describing the absence of phase information is commonly referred to by protein crystallographers as "The phase problem".

#### 6.2.3.5 Obtaining Phase Information

To drive the process of structure determination, multiple experimental approaches have been developed to extract the necessary phase information from diffracted x-rays recorded on the diffraction pattern. The most commonly employed of these approaches is the method of molecular replacement, which involves obtaining the missing phase information from a known structural homologue. In summary, the methodology underpinning molecular replacement relies on the utilisation of phase models generated from the known structural homologue and superimposition of these phases onto the unknown structure. The success of molecular replacement is thus entirely dependent on the degree of corroboration

between the sequence identities of the known structural homologue and the unknown structure. In most cases, higher levels of corroborative sequence identities correlate with greater superimposition and generate a more accurate prediction of phase information for the unknown structure. Once the structures of the known homologue and the unknown target protein have been superimposed, the resulting phase information can then be harnessed to provide a near complete derivation of the structure.

Although the method of molecular replacement has been used for many years to successfully exploit a multitude of protein structures, the employment of such an approach is not universally applicable to all protein structures. For example, proteins exhibiting novel domain architectures or unique folds for which phase information is unavailable (due to the absence of structural homologues) must derive the necessary phase information experimentally. One of the most commonly used techniques for determining phase information experimentally is known as Multiple Isomorphous Replacement (MIR), which relies on the soaking of protein crystals in a solution containing heavy metal atoms. The theoretical basis of MIR relies on the fact that heavy metal atoms are electron rich and therefore possess higher atomic scattering factors. This intrinsic property of heavy metal atoms is manifested in the production of x-ray diffraction patterns of altered intensity compared with non-heavy metal soaked versions of the same crystal. The difference in intensities of the emitted x-ray beams is sufficient to derive the experimental phase information crucial for structure determination. Heavy metal atoms typically used for crystal soaking as part of MIR include platinum, gold and iodine which all react with histidine and methionine residues, and metals from the lanthanide and actinide series which are reactive towards glutamic acid and aspartic acid residues. In order for MIR to work successfully, the heavy-atom soaked crystals must retain the same unit cell

dimensions as the native non-soaked crystals. Since x-ray diffraction data is collected on the heavy atom-soaked crystals and then subtracted from the x-ray diffraction patterns obtained from native crystals, any perturbations or alterations in the unit cell dimensions of the heavy-atom soaked crystals will result in inaccurate determination of the heavy atom positions and thus incorrect determination of the experimental phases. In order to provide independent confirmation of phase information, orthogonal heavy-atom soaking experiments are usually conducted with a second heavy metal atom. Contrary to molecular replacement approaches for obtaining phase information, MIR provides a useful approach for the experimental determination of phase information for novel protein structures, however these experiments are often time consuming are soaking experiments are not always successful.

A third approach often employed to derive experimental phase information is known as multi-wavelength anomalous diffraction (MAD). Briefly, this method operates by recording anomalous diffraction measurements at different x-ray energies where the anomalous scattering factors of heavy metal atoms significantly differ from one another. The larger changes in anomalous scattering factors are typically observed near the absorption edges and thus provide useful information regarding the derivation of experimental phases to aid structure determination. Implementation of MAD is usually facilitated by incorporation of selenium through the expression of seleno-methionine-labelled proteins which are crystallised, bombarded with high intensity x-rays and the x-ray diffraction data processed in order to determine the experimental phase information.

#### 6.2.3.6 Model building and structure refinement

Experimental determination of the phase characteristics of electron waves combined with the corresponding amplitude enables the first round of model building by calculation of electron density maps. The electron density maps define the distribution boundaries of the electron clouds around each individual atom in the protein structure. These electron clouds are then used as a template and are then interpreted by substitution of a set of atomic co-ordinates in order to fit and build the backbone and side chains of specific amino acid residues. Once a preliminary structure has been built, the model then undergoes an iterative cycle of refinement using a series of computer programmes in order to produce the best possible fit of the experimental phases to the electron density maps. Higher resolution crystal structures will generate clearer electron density maps where the atomic co-ordinates can be fitted with greater accuracy.

### 6.2.4 Crystallisation of SHP2 catalytic domain by malate screening

A detailed survey of the crystallisation literature for SHP2 catalytic domain was conducted prior to setting up crystal trials. This identified the initial conditions supporting the crystallisation of the same SHP2 catalytic domain construct used in this investigation as 20% (v/v) PEG 3350, 150 mM sodium malate pH 7.5 at 298 K using sitting drop vapour diffusion [38]. Under these conditions, a single malate ion was observed as the primary occupant of the phosphatase active site, presumably acting as a key mediator for maintaining an open and accessible crystal form, with the WPD loop in the open conformation. Although representative of an open crystal form, the structure raised initial concerns over the presence of the bound malate ion, which was deemed likely to compromise the soakability of open crystals grown under these conditions. Furthermore, at

a concentration of 150 mM, the substantial molar excess of malate would be at saturating concentrations and thus outcompeting bound malate with low affinity fragments would pose a significant challenge for determining fragment-bound structures and may prove ultimately prohibitive to fragments binding in the same location.

Unfortunately, attempts to elucidate the binding affinity of malate for the SHP2 catalytic domain by SPR were unsuccessful. However, it was suggested that optimisation of the crystallisation conditions with lower malate concentrations may aid the generation of apo-crystals required for fragment soaking experiments. In addition, the decision to proceed with optimising crystallisation conditions at lower malate concentrations was also motivated by unsuccessful attempts to replicate the growth of SHP2 crystals under literature reported conditions of 20% (v/v) PEG 3350, 150 mM sodium malate pH 7.5 (Figure 6.2A) that were morphologically amenable for x-ray diffraction studies.

Hence, to investigate whether growth of SHP2 catalytic crystals was feasible at lower malate concentrations and explore the effect of this phenomenon on crystal size and morphology, a malate screening strategy was employed at different concentrations of SHP2 catalytic domain (17 mg/ml, 25 mg/ml and 34 mg/ml) using a malate concentration gradient from the highest concentration of 150 mM to the lowest concentration of 10 mM. This strategy enabled screening over a range of malate concentrations, including concentrations below 150 mM malate reported in the original conditions, and provided a rapid assessment of whether these conditions were amenable for producing crystals of appropriate morphology for x-ray diffraction. As a further consideration, it was reasoned that any reduction in malate concentration may also aid in generating the large supply of apo-crystals necessary for fragment soaking experiments. Finally, when considering the

putative importance of the malate ion for holding the active site in an open conformation, an additional area of uncertainty centred on the consequences that reducing the malate concentration may have on the WPD loop conformation and thus the overall accessibility of the phosphatase active site for fragment soaking.

The results from the small-scale malate crystallisation screens suggested that it was possible to achieve ubiquitous micro-crystal growth of SHP2 catalytic crystals across all malate concentrations and at protein concentrations of 17 mg/ml, 25 mg/ml and 34 mg/ml. Notably, clear improvements in crystal size and morphology were observed in drops containing 75 mM sodium malate pH 7.5 concentrations and above, across all protein concentrations screened. In an attempt to optimise crystal growth, large scale sitting and hanging drop crystallisation trials were set up with conditions containing protein concentrations of 17 mg/ml, 25 mg/ml and 34 mg/ml at malate concentrations of 75 mM, 100 mM, 125 mM and 150 mM. Unfortunately, despite scaling up drop sizes from 200 nl (100 nl protein: 100 nl precipitant used in the initial screens) to 4 µl (2 µl protein: 2 µl precipitant), no clear improvement in crystal size and morphology was observed. However, closer inspection of crystal conditions containing 75 mM sodium malate pH 7.5 at the lowest protein concentration of 17 mg/ml (Figure 6.2B) revealed that there were minor improvements in the size and morphology of crystals grown in this condition.

Despite observing this improvement in crystal growth, the abundance of crystal growth in each drop suggested that a reduction in the number of nucleation sites in each drop was a worthwhile strategy for further optimisation of crystal morphology.

However, reducing the protein and precipitant concentration by setting up crystallisation conditions containing 14 mg/ml - 17 mg/ml protein and 15%-20% (v/v) PEG 3350 only

resulted in minor improvement in crystal morphology and micro-crystal growth (Figure 6.2C). This suggested a more drastic adjustment of the crystallisation conditions were necessary in order to achieve sufficient reduction in the levels of nucleation and optimisation of crystal morphology.

### 6.2.5 Optimisation of SHP2 catalytic domain crystal morphology

To further optimise crystal morphology large-scale crystal screens were set up at lower protein, malate and precipitant concentrations in order to reduce the number of nucleation sites and achieve crystal morphologies representative of diffraction-quality crystals. Interestingly, screening at reduced protein concentrations of 8 mg/ml and 10 mg/ml using malate concentrations of 10 mM and 25 mM and a PEG 3350 concentration range of 7.5%-10% (v/v) produced significantly better crystal morphology compared to results obtained from previous crystal trials. In particular, large scale optimisation of these conditions resulted in distinct morphological improvement of crystals grown under conditions containing 8 mg/ml protein concentration at 8% (v/v) PEG 3350 with 10 mM and 25 mM sodium malate pH 7.5 in the crystallisation buffer (Figure 6.2D).

### 6.2.6 Crystallisation of SHP2 catalytic domain in the presence of DMSO

To explore whether crystals exhibited the robust properties required for fragment soaking studies and confirm whether crystal integrity was preserved in the presence of DMSO, crystals were grown under the optimised condition containing 8 mg/ml protein concentration, 10 mM and 25 mM sodium malate pH 7.5, 8% (v/v) PEG 3350, in the presence of 5% (v/v) DMSO. Strikingly, despite the addition of 5% (v/v) DMSO these conditions yielded crystals with excellent morphology with no obvious defects in crystal appearance and integrity, suggesting that crystals grown under these conditions exhibited

DMSO tolerance and could be produced under the optimised crystal conditions in both the presence and absence of 5% (v/v) DMSO. The ability of these crystals to withstand harsh crystallisation conditions containing 5% (v/v) DMSO was also reassuring suggesting that crystal integrity would likely be preserved during fragment soaking experiments conducted at high DMSO concentrations.

After an iterative cycle of crystal optimisation and assessing crystal growth across a range of conditions, a number of crystals were identified from a large scale screen of the optimised crystal condition containing 8 mg/ml protein concentration in conditions containing 8%-9% (v/v) PEG 3350, 25 mM sodium malate pH 7.5 grown in the presence and absence of 5% (v/v) DMSO, with drop sizes of 1 µl: 2 µl and 1 µl to 3 µl protein: precipitant respectively. These crystals were selected for x-ray diffraction analysis on the basis of size and depth with larger crystals exhibiting straighter edges used as the predominant criterion for optimum morphology (Figure 6.2E).

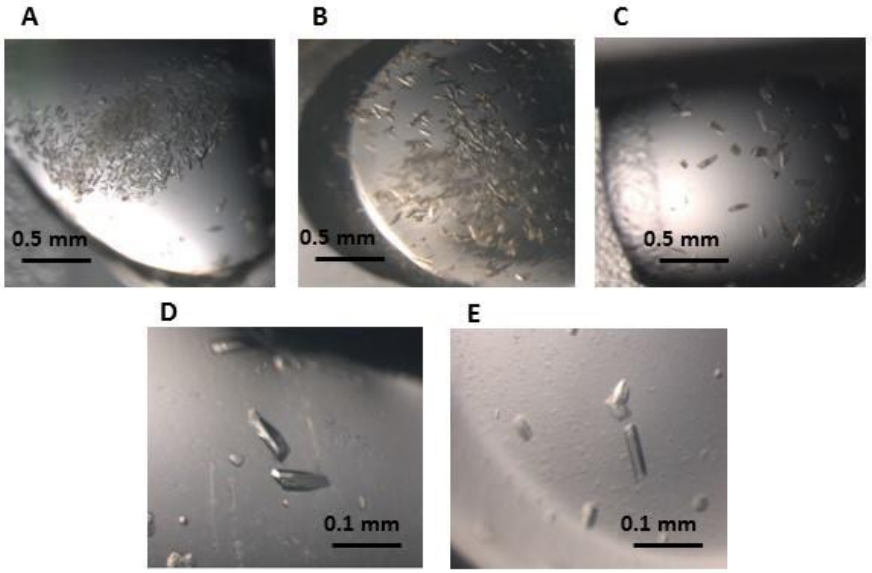



**Figure 6.2. Small and large-scale optimisation of SHP2 catalytic domain crystals.** A) Small scale microcrystals of SHP2 catalytic domain grown in conditions containing 25 mg/ml protein concentration, 20% (v/v) PEG 3350, 150 mM sodium malate pH 7.5, with 100 nl: 100 nl protein: precipitant set up using a Mosquito robot. B) Small scale growth of SHP2 catalytic domain crystals in conditions containing 20% (v/v) PEG 3350, 75 mM sodium malate pH 7.5 set up at protein concentrations of 17 mg/ml. C) Reduction of crystal nucleation sites by setting up conditions at reduced protein and precipitant concentrations containing 14 mg/ml-17 mg/ml protein and 15%-20% (v/v) PEG 3350 only resulted in minor improvement in crystal morphology and micro-crystal growth. D) Large scale optimisation of crystallisation conditions containing 8 mg/ml protein concentration in a crystallisation buffer containing 25 mM sodium malate pH 7.5 and a PEG 3350 concentration range of 7.5%-10%. These conditions generated diffraction-quality crystals which were subsequently used to derive the structure of the SHP2 catalytic domain in the apo-conformation. E) Large scale optimisation of DMSO-tolerant SHP2 catalytic domain crystals grown at a protein concentration of 8 mg/ml and under crystallisation conditions containing 8%-9% (v/v) PEG 3350 and 25 mM sodium malate pH 7.5, grown in the presence of 5% (v/v) DMSO.

### 6.2.7 Data collection and processing

X-ray diffraction experiments were performed as described in Material and Methods (see section 2.9). Briefly, prior to x-ray data collection, crystals grown at 8 mg/ml protein concentration in crystallisation buffer containing 8% (v/v) PEG 3350 and 25 mM sodium malate pH 7.5 were soaked in a cryoprotectant mixture containing 25% (v/v) ethylene glycol as previously described [38]. After cryoprotection, crystals were then mounted and cooled at 100 K under a nitrogen gas cryostream. X-ray diffraction data was collected to a resolution of 3.1 Å (Figure 6.3) by Dr Fiyaz Mohammed on an in-house MicroMax 007HF X-ray source (Rigaku) using a Saturn CCD detector (X-ray facility, School of Biosciences, University of Birmingham). The initial data collected was auto indexed with XDS and revealed that the SHP2 catalytic domain crystallised in the orthorhombic space group P21 with unit cell parameters a = 48.7 Å, b = 130.2 Å, c = 48.9 Å, α = 90o, β= 109.6$^{o}$ and γ = 90$^{o}$. The raw X-ray diffraction data was then integrated, scaled and merged using programs of the XDS suite. The relevant data processing statistics are displayed in table 6.1.

| | |
|---|---|
| Space Group | $P2_1$ |
| Unit Cell Length (Å) | **a= 48.7 Å, b = 130.2 Å, c = 48.9 Å** |
| Unit Cell Angles ($^O$) | **$\alpha = 90^o$, $\beta = 109.6^o$, $\gamma = 90^o$** |
| Resolution range (Å) | **65-3.1** |
| Total Reflections | **147143 (13248)** |
| Unique Reflections | **10352 (929)** |
| Completeness (%) | **98.5 (97.7)** |
| $R_{sym}$ (%) | **15.9 (60.9)** |
| $R_{free}$ (%) | **20.2** |
| I/σ (I) | **20.99 (5.6)** |
| Multiplicity | **1.37 (14.27)** |

**Table 6.1 Data processing statistics for structure determination of SHP2 catalytic domain**. X-ray data processing statistics for structure determination of SHP2 catalytic domain were obtained by integrating, scaling and merging using XDS. Numbers in parentheses represent data from the highest resolution shell.

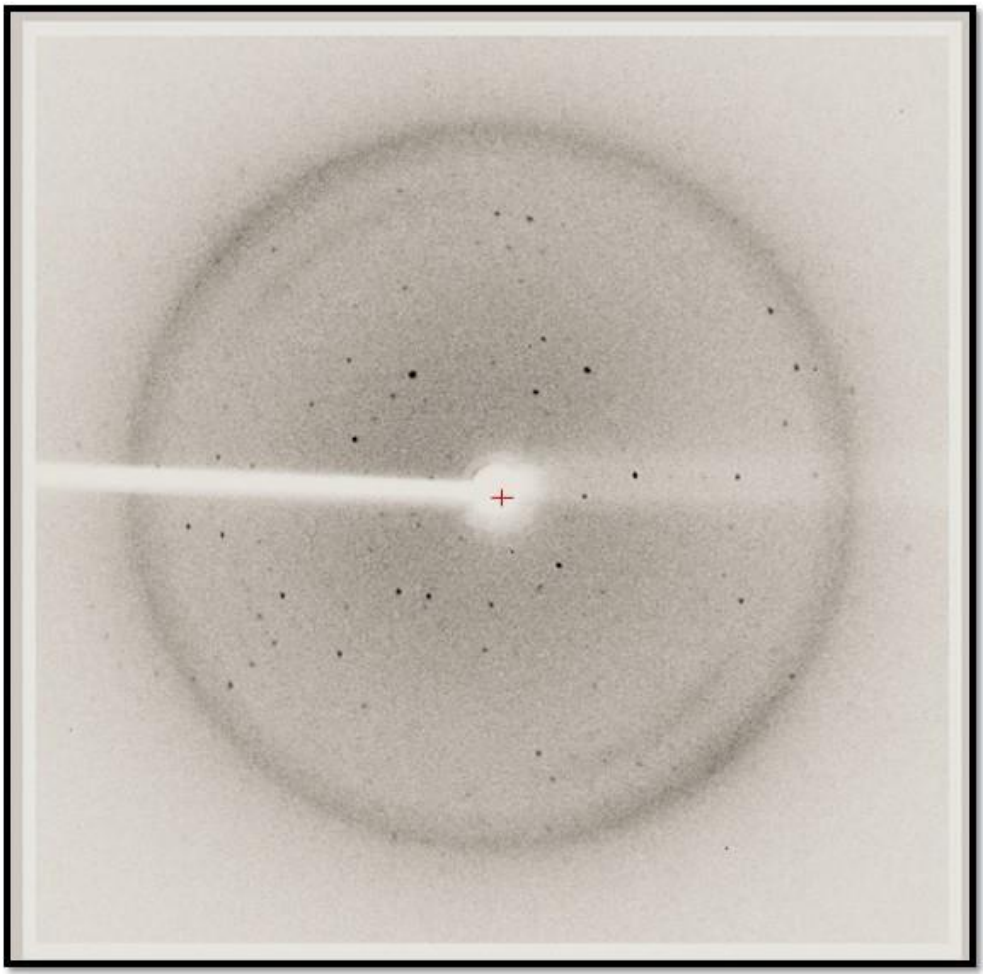


**Figure 6.3 X-ray diffraction pattern of SHP2 catalytic domain**. A 0.5$^{o}$ oscillation image of the SHP2 catalytic domain crystal recorded on an in house Micro-Max 007HF rotating anode X-ray generator (Rigaku). The crystal generated an X-ray diffraction pattern to a resolution of 3.1Å.

### 6.2.8 Structure determination using molecular replacement

In order to obtain the necessary phase information for structure determination of the SHP2 catalytic domain, the method of molecular replacement was employed using MOLREP and utilising the previously determined malate-bound structure of the SHP2 catalytic domain (PDB: 3B7O) as the primary search model. The criteria for selection of a suitable search model was based solely on the very high degree of sequence identity (100%) exhibited by the experimentally derived SHP2 structure and 3B7O, both of which were constructs of identical sequence. Unsurprisingly, the outputs from molecular replacement calculations yielded unambiguous solutions for the translation and rotation functions (Tables 6.2, 6.3 and 6.4) consistent with the SHP2 catalytic domain crystallising as a dimer with two copies in the asymmetric unit. After an initial round of refinement, the R-factors of the resulting model were reduced from 0.523 to 0.328 (Table 6.4).

| Peak number | α | β | γ | $R_f/\theta$ |
|---|---|---|---|---|
| 1 | **5.00** | **68.59** | **240.39** | **12.38** |
| 2 | **266.27** | **13.49** | **336.51** | **12.38** |
| 3 | **0.00** | **66.98** | **230.82** | **4.97** |
| 4 | **124.28** | **33.31** | **207.25** | **4.17** |
| 5 | **329.83** | **54.55** | **164.47** | **4.14** |

**Table 6.2 Rotation function statistics for structure determination of SHP2 catalytic domain using molecular replacement**. The table shows the outputs from molecular replacement which contains the 5 highest solutions for the rotation function peaks using structure 3B7O as the primary search model. The correct solutions for the rotation function peaks are highlighted in red, which were determined by the increase in Rf/θ score.

| Peak number | $T_X$ | $T_Y$ | $T_Z$ | $RF_{cnt}$ | R factor |
|---|---|---|---|---|---|
| 1 | **0.201** | **0.00** | **0.159** | **21.52** | **0.478** |
| 2 | **0.352** | **0.00** | **0.252** | **18.95** | **0.488** |
| 3 | **0.207** | **0.00** | **0.156** | **2.95** | **0.550** |
| 4 | **0.463** | **0.00** | **0.465** | **2.79** | **0.561** |
| 5 | **0.317** | **0.00** | **0.364** | **2.38** | **0.557** |

**Table 6.3 Translation function statistics for SHP2 catalytic domain (PDB code 3B7O).** The table shows the 5 highest translation function solutions determined using the rotation function angles derived from Table 6.2. The correct solutions for the translation function peaks are determined by a progressive decrease in the R factor value.

| Peak number | $T_X$ | $T_Y$ | $T_Z$ | $RF_{cnt}$ | R factor |
|---|---|---|---|---|---|
| 1 | **0.352** | **0.238** | **0.752** | **31.22** | **0.368** |
| 2 | **0.209** | **0.355** | **0.155** | **1.22** | **0.525** |
| 3 | **0.768** | **0.804** | **0.849** | **2.81** | **0.524** |
| 4 | **0.342** | **0.700** | **0.960** | **3.51** | **0.525** |
| 5 | **0.124** | **0.817** | **0.720** | **3.11** | **0.523** |

**Table 6.4 Translation function statistics for SHP2catapo**. The table shows the 5 highest translation function peaks after fixing the search model structure (PDB: 3B7O). The correct solutions are highlighted in red which are determined by a progressive decrease in the R factor value.

After calculation and interpretation of the relevant statistical outputs, the molecular replacement solution was further validated by assessment of the calculated electron density maps. Unsurprisingly, the identical sequences for the SHP2 catalytic domain and the search model 3B7O generated clear and unambiguous electron density allowing the structure determination with a single round of refinement. Typically, fitting of the experimentally determined structure to the calculated electron density maps is followed by iterative cycles of model building and structure refinement in order to generate the final refinement and stereochemical restraints necessary for complete structure determination. However, for the purposes of this investigation, further cycles of model building and refinement were deemed unnecessary as a single round of refinement was sufficient to confirm the structure, given the already available structure of 3B7O with identical sequence.

### 6.2.9 Fragment and inhibitor soaking of SHP2 catalytic domain crystals

In addition, the crystal of the SHP2 catalytic domain solved to 3.1Å was soaked in a cryoprotectant solution containing the SHP2 catalytic domain inhibitor NSC-87877. Unfortunately, we were unable to identify any clear electron density to unambiguously define the binding site of this inhibitor. Similar efforts to derive fragment-bound structures of the SHP2 catalytic domain also resulted in similar outcomes with no observable electron density detected in the final refined structure.

### 6.2.10 Structural analysis of the SHP2 catalytic domain crystal structure

In the absence of further structural refinement the experimentally derived crystal structure of the SHP2 catalytic domain (Figure 6.4) is expected to be comparable to previously determined structures solved in both malate-bound (PDB: 3B7O) and auto-inhibited (PDB: 2SHP) conformations. However, the lower resolution of this structure compared with the malate-bound structure (which was refined to a final resolution of 1.8 Å) prohibits further discussion regarding the presence or absence of bound ligands or crystallisation artefacts, and greatly limits the conclusions that can be drawn regarding whether the catalytic site is in an open or closed conformation.

Structurally, the architecture of the SHP2 catalytic domain comprises a total of 8 α helices and 13 β strands which is topologically identical to structure 3B7O, but is structurally different from 2SHP which features an additional β strand and α helix corresponding to N-terminal amino acid residues 221-234 which are truncated from the 3B7O construct. Ten of the thirteen β strands group to form a large β sheet dominated by a combination of parallel and anti-parallel architecture which encloses the isolated helix αE. Helices αB, αF, αG, αH, αI and αJ cluster together on the opposite face of the dominant β sheet. The

signature catalytic motif containing the sequence [I/V]HCXAGXXR[S/T]G (where X denotes any amino acid) which harbours the catalytically conserved nucleophilic cysteine amongst other key functional groups required for phosphate recognition are located in sheet βM, helix αG and the intermittent loop region connecting these two elements. The catalytic motif in conjunction with the essential WPD loop (the latter formed from the loop region extending from sheet βL and terminating at the start of helix αF) form a catalytic groove at the base of the phosphate binding pocket.

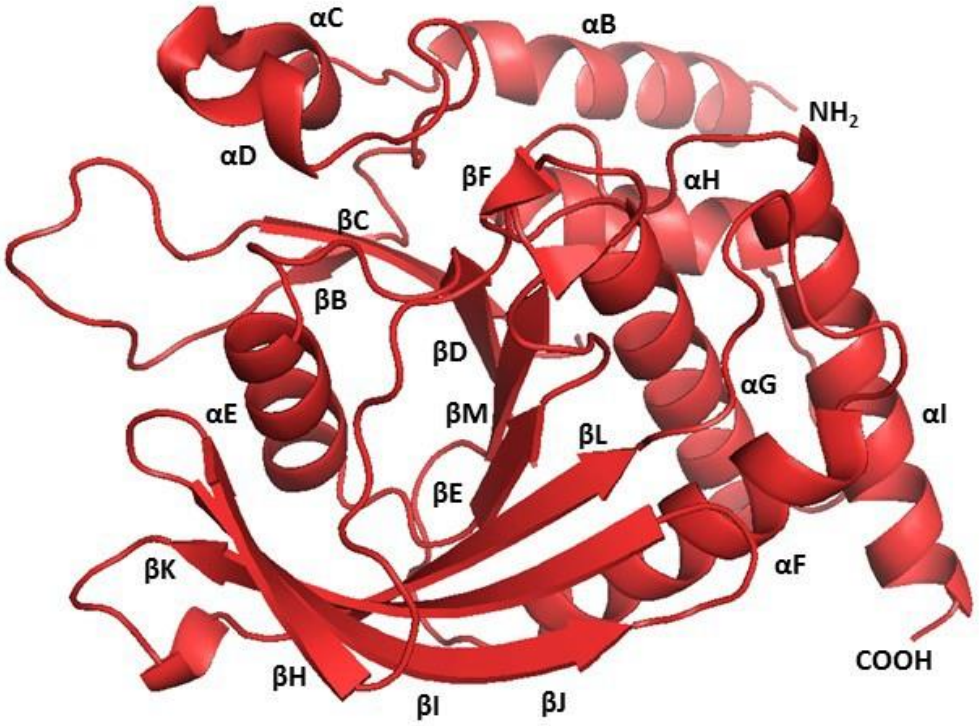



**Figure 6.4. Crystal structure of the SHP2 catalytic domain**. The experimentally determined crystal structure of the SHP2 catalytic domain was solved by molecular replacement using the homologue structure 3B7O. Assignment of individual secondary structure motifs comprising the mixed α/β architecture was based on the previously reported sequence alignments of the SHP2 catalytic domain [80].

### 6.2.11 Crystal structure of the SHP2 catalytic domain dimer

Since two molecules of SHP2 catalytic domain were found to occupy the asymmetric unit (Figure 6.5), the experimentally derived crystal structure of the SHP2 catalytic domain initially raised the question of whether the dimeric interface captured during crystallisation was a physiologically representative state of the SHP2 catalytic domain or merely an artefact of crystal packing. Interestingly, a similar dimeric interaction in the asymmetric unit was also observed during crystallisation of the SHP2 catalytic domain in the auto-inhibited conformation (PDB: 2SHP). Therefore, to confirm the physiological relevance of the dimer, the residues stabilising the dimer interface of both SHP2catapo and 2SHP were analysed in further detail to observe whether these were conserved across both structures. If dimerization of SHP2 was functionally significant, then one might expect residues governing this interaction to have been conserved throughout evolution to support the signalling capabilities of SHP2. Indeed, analysis of the residues occupying the dimer interface of the SHP2 catalytic domain reveal that the majority of the stabilising interactions form between the loop region of beta sheet βK of molecule A with the corresponding loop region of beta sheet βF of molecule B. Additional stabilising interactions are also predicted between the loop region of beta sheet βK from molecule A and the helix of beta sheet βF of molecule B, and the loop region of alpha helix αF in molecule A with the loop regions of beta sheets βL and βE in molecule B (Table 6.5). Similar analysis of the dimeric interface stabilising the crystal structure of 2SHP (Figure 6.6) also revealed that the majority of the residues lining the dimer interface are located in the loop region of beta sheets βJ and βL in molecule A and the loop region of beta sheet βF in molecule B. In particular, the loop region of beta sheet βF was also identified as a key interaction motif stabilising the dimer interface. Despite a commonality between

interaction motifs predicted to stabilise the dimer interface reported in the experimentally derived crystal structure of SHP2 and auto-inhibited crystal structure of the SHP2 catalytic domain, the evidence gathered in this study does not support the physiological relevance of this dimeric interaction. In particular, the low resolution of the experimentally derived crystal structure of the SHP2 catalytic domain suggest that these predictions regarding the residue contacts between the dimer interface must be treated with caution. Further evidence against the notion of a physiologically relevant dimer of the SHP2 catalytic domain reported in this study includes size-exclusion chromatography, which revealed that the SHP2 catalytic domain elutes with a single dominant peak indicative of a dominant monomeric species, and that this observation is identical in both physiologically relevant phosphatase buffer conditions as well as and non-physiologically relevant HEPES buffer conditions. This is also supported by AUC data which demonstrates that SHP2 also exists as a predominantly monomeric species in solution, in both phosphate and HEPES buffer conditions (see section 3.2.6).

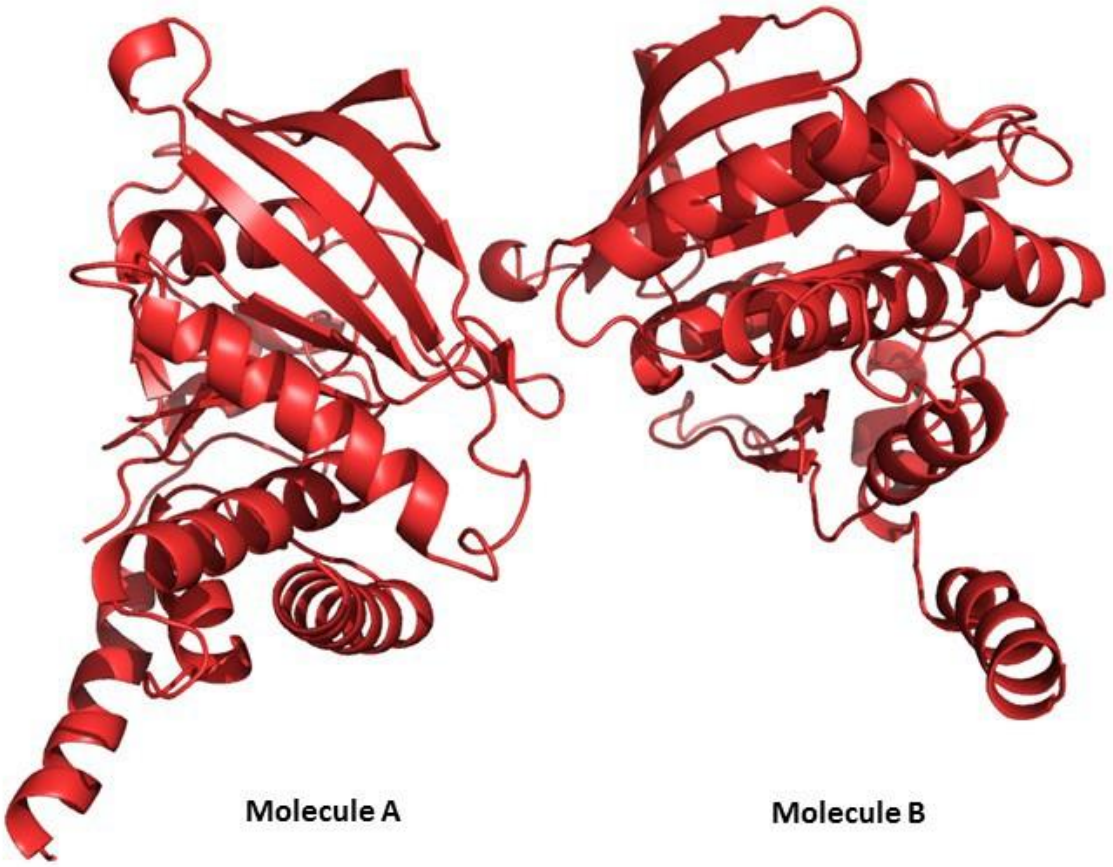



**Figure 6.5 Crystal structure of the SHP2catapo dimer**. The experimentally determined crystal structure of SHP2 was observed to contain two molecules in the asymmetric unit (molecule A and molecule B).

| Molecule A | Location | Molecule B | Location | Distance |
|---|---|---|---|---|
| Q408 | βK (L) | G363 | βF (L) | 3.15 Å |
| N410 | βK (L) | S365 | βF (L) | 2.51 Å |
| N410 | βK (L) | V360 | βF (H) | 2.86 Å |
| T411 | βK (L) | V360 | βF (H) | 3.41 Å |
| M450 | αF (L) | R421 | βL (L) | 2.54 Å |
| E451 | αF (L) | K358 | βE (L) | 3.03 Å |

**Table 6.5 Residues involved in stabilisation of the SHP2 catalytic domain dimer interface**. The majority of the residues occupying the interface are located in the loop region (L) of beta sheet βK of molecule A with the corresponding loop region (L) of beta sheet βF of molecule B. Additional residues include the loop region of alpha helix αF in molecule A with the beta sheets of βL and βE. Distances between the interacting side chains of these residues are also included.

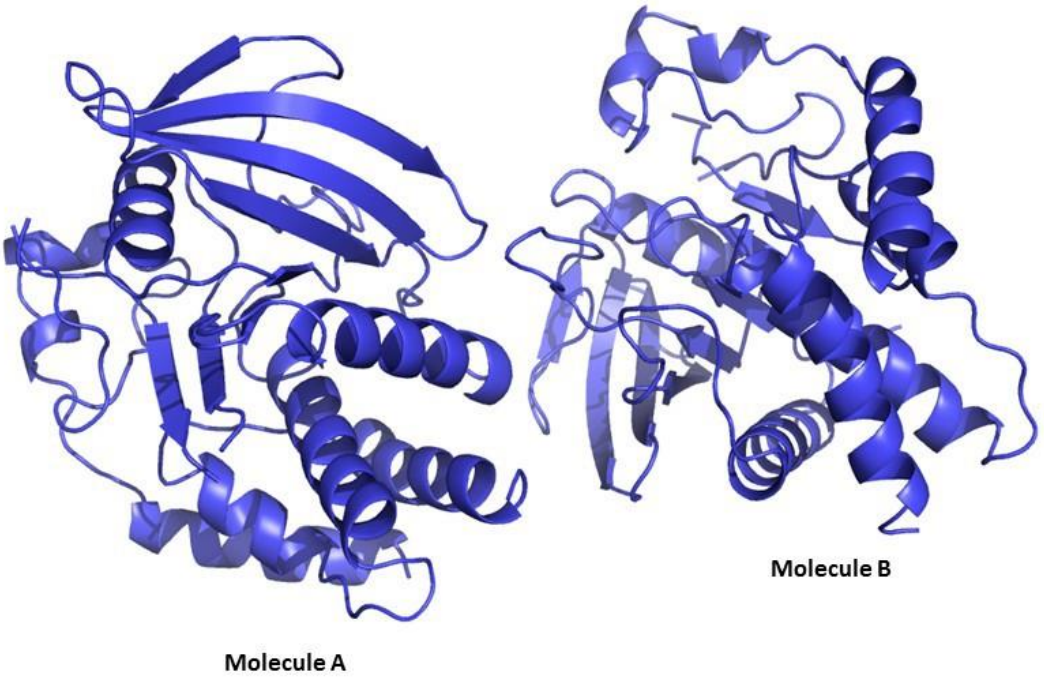



**Figure 6.6 Crystal structure of the auto-inhibited SHP2 catalytic domain dimer from full length SHP2 (PDB: 2SHP).** Structure determination of auto-inhibited full length SHP2 observed two copies of each molecule in the asymmetric unit, molecule A and molecule B.

| Molecule A | Location | Molecule B | Location | Distance |
|---|---|---|---|---|
| H394 | βJ (L) | G363 | βF (L) | 2.65 Å |
| H394 | βJ (L) | S365 | βF (L) | 3.18 Å |
| S430 | βL (L) | D373 | βF (L) | 2.86 Å |

**Table 6.6 Residues involved in stabilisation of the auto-inhibited SHP2 catalytic domain dimer interface**. The key residues occupying the dimer interface are located in the loop region (L) of beta sheet βJ of molecule A with loop region (L) of beta sheet βF of molecule B, and loop region (L) of beta sheet βL on molecule A with loop region (L) of beta sheet βF. Distances between interacting side chains of these residues are also included.

### 6.2.12 Superimposition of SHP2 catalytic domain with PDB structures 3B7O and 2SHP

In order to confirm whether the experimentally derived structure of the SHP2 catalytic domain was structurally similar or exhibited any structural perturbations compared with the previously determined malate-bound (PDB: 3B7O) and auto-inhibited conformations of the SHP2 catalytic domain (PDB: 2SHP), all three molecules were superimposed and the Root Means Square Deviation (RMSD) calculated (Figure 6.7). As predicted, an RMSD of 0.17 Å was calculated for the experimentally derived SHP2 crystal structure and the malate-bound SHP2 catalytic domain (PDB:3B7O), suggesting that these molecules possess very high structural similarity. However, the RMSD values for superimposition of 2SHP onto SHP2 catalytic domain yielded a value of 3.8 Å, which was likely a result of the profoundly distorted N-terminal helix in 2SHP which extends into the C-SH2 and N-

SH2 domains of SHP2 to form the entire full length SHP2 protein. The distorted helix geometry is therefore likely to resemble a physiologically relevant conformation as positioning of this helix is crucial for orienting the tan-SH2 domain in the correct position to mediate occlusion of the catalytic pocket and maintain the integrity of the auto-inhibited state.

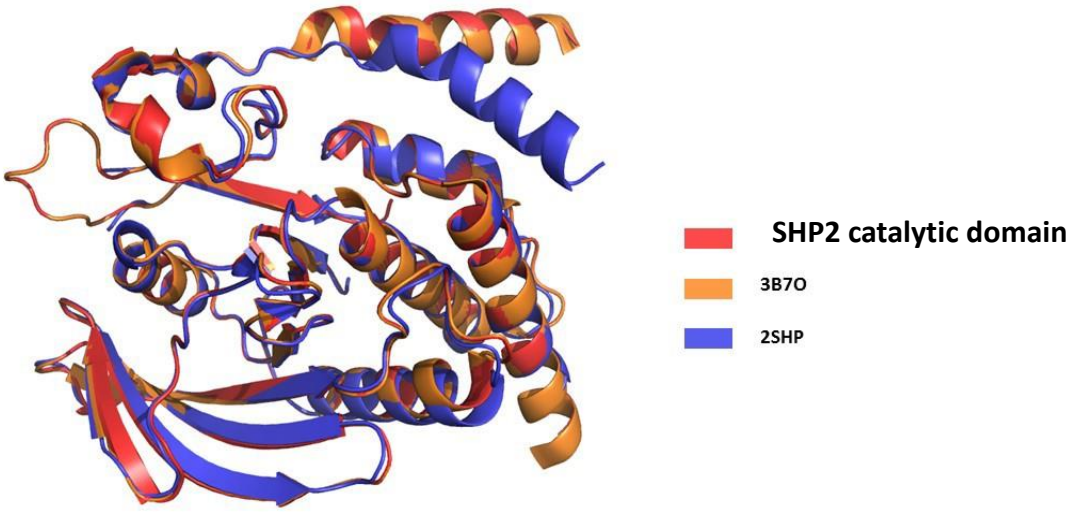



**Figure 6.7 Overlay of superimposed structures of the experimentally derived crystal structure of SHP2 with PDB structures 3B7O and 2SHP**. Crystal structures of SHP2 catalytic domain (red), 3B7O (orange) and 2SHP (blue) were all superimposed to enable RMSD calculations. Superimposition of SHP2catalytic domain and 3B7O both gave an RMSD value of 0.17 Å, signifying a high degree of structural similarity. The RMSD value for 2SHP superimposed to SHP2 catalytic domain and 3B7O was 3.8 Å, presumably owing to the distorted helix geometry which extends to the tan-SH2 domain of SHP2 and functions by maintaining the integrity of the auto-inhibited state of the full length SHP2 protein.

### 6.2.13 WPD loop conformation in the SHP2 catalytic domain crystal structure

Although a crystal structure of the SHP2 catalytic domain was solved to a resolution of 3.1Å, no definitive conclusions can be drawn regarding the presence of bound malate despite crystals being grown in conditions containing high malate concentrations. Given this information, we were intrigued to observe the conformational state of the critical phosphatase WPD loop in the experimentally derived crystal structure of SHP2. In addition to playing a central role in the phosphatase catalytic mechanism, the WPD loop is a critical regulator of substrate access and therefore acts as a gateway by governing the accessibility of the phosphatase active site to a variety of phosphopeptide ligands [33]. The malate bound structure of SHP2 catalytic domain (PDB: 3B7O) reports the WPD loop in an open and accessible conformation, suggesting that the presence of the malate ion is acting to support an open conformation of the active site. Initially, although the objective was to generate a structure of the SHP2catalytic domain in the absence of bound malate, it was unclear whether generating a structure of this kind would destabilise the WPD loop thus leading to a closed and inaccessible active site. However, analysis of the electron density map for the WPD loop region (Figure 6.8B) suggests that, in the absence of bound malate, the WPD loop appears to occupy an open conformation with the W423 side chain extending below the plane of the catalytic cleft mimicking the W423 side chain conformation observed in the malate bound structure of the SHP2 catalytic domain (Figure 6.8). This position of the W423 side chain is in stark contrast to the conformation observed in the auto-inhibited structure 2SHP (Figure 6.8A, blue) where the upwards protrusion of the W423 side chain is sufficient to occlude the catalytic pocket. However, as previously mentioned, the low resolution of this structure compared to the malate-bound crystal

structure of SHP2 (PDB: 3B70) is insufficient to conclusively define both the conformational state of the WPD loop and the presence or absence of a bound malate ion.

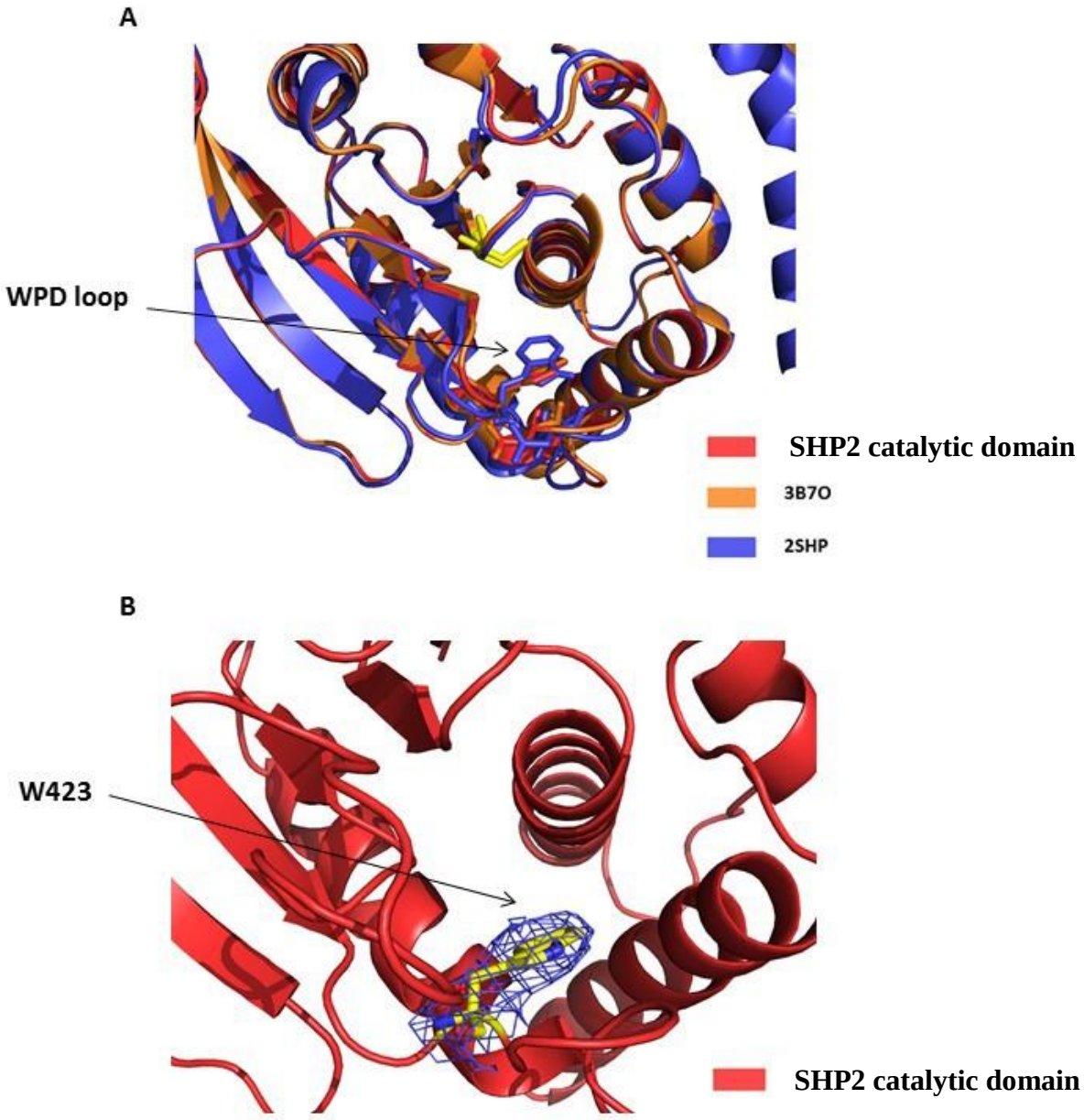



**Figure 6.8. Comparison of the WPD loop conformation in SHP2 catalytic domain, 3B7O and 2SHP**. A. Overlay of the crystal structures of SHP2 catalytic domain (red), 3B7O (orange) and 2SHP (blue). WPD loop conformations (highlighted as sticks) highlight the position of the W423 side chain as critical determinant for governing open and closed conformations of the phosphatase catalytic cleft. The catalytic cysteine is highlighted in yellow. B. Electron density of the critical W423 residue in the SHP2 catalytic domain structure suggests that the W423 side chain conformation may mimic the equivalent conformation of the identical W423 residues observed in structure 3B7O leading to an open and accessible catalytic cleft. However, the limitation of a lower resolution structure prohibits any definitive conclusions from being drawn regarding the precise conformation of the WPD loop.

## 6.3 Discussion

Fragment-based ligand discovery relies on the robust validation of initial fragment hits to accelerate the development of hit matter into potent and selective lead molecules. The evolution of lead inhibitors requires a robust model of fragments binding to the target of interest, insights of which are provided by x-ray crystallography. This enables progressive and iterative expansion of the fragment through multiple cycles of medicinal chemistry, ultimately leading to the development of more potent inhibitors. Following a fragment-based library screen of the SHP2 domain, which identified fragments 1A9 and 4B1 as active and validated fragment hits, the overarching aim of this chapter was to explore and optimise a known set of crystallisation conditions for the SHP2 catalytic domain to aid the development of an appropriate crystal form that was amenable to fragment-soaking. The appropriate crystal form could then be utilised to determine the crystal structures of fragments 1A9 and 4B1 bound to the SHP2 catalytic domain and drive optimisation and expansion of these fragment into more potent inhibitors of the SHP2 catalytic domain.

In order to develop an appropriate crystal form of the SHP2 catalytic domain, an initial set of conditions were used as a starting point for further optimisation. These crystallisation conditions contained 20% (v/v) PEG 3350 and 150 mM sodium malate pH 7.5, and had previously led to the derivation of the first crystal structure of the SHP2 catalytic domain in an open and accessible conformation [38]. Interestingly, the crystal structure reported by Barr, et al. revealed the SHP2 catalytic domain in complex with a bound malate ion, which was presumably acting as a key mediator for stabilisation of the active site in an open conformation. Although an open and accessible active site is necessary for fragment-soaking, the presence of the bound malate ion as a crystallisation artefact would be prohibitive to fragments binding in the same location. Furthermore, malate concentrations

of 150 mM (reported in the initial crystallisation conditions) would also serve to outcompete fragments binding to the active site. In order to generate a crystal form devoid of bound malate, a number of crystal conditions were screened across a malate concentration range spanning 10 mM to 150 mM, the latter malate concentration being the condition that resulted in the successful crystallisation of the malate-bound open conformation of the SHP2 catalytic domain. After several rounds of optimisation, final conditions of 8 mg/ml SHP2 catalytic domain, 8%-9% (v/v) PEG and 25 mM malate was sufficient for the generation of diffraction-quality crystals. Interestingly, crystals of similar morphology also grew under the same conditions supplemented with 5% (v/v) DMSO. Although these crystals did not diffract, crystal growth in conditions supplemented with 5% (v/v) DMSO suggest that crystals exhibited sufficient DMSO tolerance and thus would be capable of withstanding the high DMSO concentrations necessary for soaking fragments 1A9 and 4B1 into crystals of the SHP2 catalytic domain.

Despite exploring a range of different soaking conditions, including different concentrations of fragments 1A9 and 4B1 in the final soak, there was no clear electron density confirming the presence of a bound fragment or inhibitor to the SHP2 catalytic domain to validate the soakability of the crystal system. Although electron density for fragments and inhibitors was not observed in fragment and ligand-soaked crystals, additional factors may be responsible for the lack of crystal soakability. The most important of these factors which underlies the soakability of all crystal systems is the presence (or absence) of an open and accessible solvent channel in the crystal leading to an unobstructed binding site. Although there is no clear method for detecting the presence of crystal solvent channels, crystals with high levels of symmetry and open lattice structures generally correlate with excellent levels of soakability [224]. To overcome the problem of

crystal soakability, extensive engineering of the protein construct or utilisation of a surrogate form of the protein with high sequence homology have been typically employed in fragment-based screening campaigns to generate more tractable crystal systems [229].

Interestingly, despite the availability of multiple inhibitor-bound structures of the SHP2 catalytic domain [143, 148, 230], all of these have been successfully generated by the use of co-crystallisation as opposed to soaking strategies. Indeed, co-crystallisation strategies for structure determination of protein-ligand complexes possess several advantages over fragment and ligand-based soaking methods. For example, co-crystallisation typically generates a more accurate representation of native ligand binding modes as incubation of the protein and ligand under solution conditions enables a greater range of dynamic and conformational binding equilibria to be sampled, compared with the relatively static dynamics of a crystal structure.

Despite unsuccessful attempts to generate structures of fragments 1A9 and 4B1 bound to the SHP2 catalytic domain, crystal screening enabled the determination of the structure of the SHP2 catalytic domain crystallised using an alternative set of crystallisation conditions. However, the low resolution of the final crystallisation structure coupled with only a single round of structural refinement prohibit drawing definitive conclusions regarding the presence or absence of bound ligands. The same degree of caution must also be utilised in the interpretation of the crystal dimer and the residue contacts governing the dimeric interface.

## 6.4 Conclusion

After conducting a fragment-based library screen against the SHP2 catalytic domain, which identified fragments 1A9 and 4B1 as validated fragment hits, the work conducted in this chapter centred on the exploration and optimisation of a known set of crystallisation conditions to generate an alternative crystal form of the SHP2 catalytic domain amenable for structure determination of the SHP2 catalytic domain bound with fragment 1A9 and 4B1. Unfortunately, although efforts to derive fragment-bound structures were unsuccessful, a crystal structure of the SHP2 catalytic domain was solved with a single round of refinement to a resolution of 3.1Å. Although the structural features of this structure were comparable with the malate-bound structure 3B7O, any conclusion regarding the presence of bound malate must be treated with caution.

Despite the amenability of phosphatase enzymes to fragment-based ligand discovery, the inherent challenges associated with the development of selective phosphatase inhibitors have meant that soakable crystal structures for phosphatases have been a relatively underexplored area of research. With fragment-based screening now widely established as a powerful tool for identifying novel binding sites and ligandable pockets, particularly in the context of more challenging target classes, future applications of fragment-based drug discovery may lead to a surge in the pipeline of soakable phosphatase crystal structures to facilitate structure-based discovery efforts against this notoriously challenging class of enzymes.

## References


1. Hunter, T., *Signaling--2000 and beyond.* **Cell**, 2000. 100(1): p. 113-27.

2. Pawson, T. and J.D. Scott, *Protein phosphorylation in signaling--50 years and counting.* **Trends Biochem Sci**, 2005. 30(6): p. 286-90.

3. Moorhead, G.B., V. De Wever, G. Templeton, and D. Kerk, *Evolution of protein phosphatases in plants and animals.* **Biochem J**, 2009. 417(2): p. 401-9.

4. Feher, Z. and K. Szirak, *Signal transduction in fungi--the role of protein phosphorylation.* **Acta Microbiol Immunol Hung**, 1999. 46(2-3): p. 269-71.

5. Deutscher, J. and M.H. Saier, Jr., *Ser/Thr/Tyr protein phosphorylation in bacteria - for long time neglected, now well established.* **J Mol Microbiol Biotechnol**, 2005. 9(3-4): p. 125-31.

6. Luan, S., *Protein phosphatases in plants.* **Annu Rev Plant Biol**, 2003. 54: p. 63-92.

7. Olsen, J.V., B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, and M. Mann, *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* **Cell**, 2006. 127(3): p. 635-48.

8. Hunter, T. and B.M. Sefton, *Transforming gene product of Rous sarcoma virus phosphorylates tyrosine.* **Proc Natl Acad Sci U S A**, 1980. 77(3): p. 1311-5.

9. Hunter, T., *The role of tyrosine phosphorylation in cell growth and disease.* **Harvey Lect**, 1998. 94: p. 81-119.

10. Alonso, A., J. Sasin, N. Bottini, I. Friedberg, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, and T. Mustelin, *Protein tyrosine phosphatases in the human genome.* **Cell**, 2004. 117(6): p. 699-711.

11. Larsen, M., M.L. Tremblay, and K.M. Yamada, *Phosphatases in cell-matrix adhesion and migration.* **Nat Rev Mol Cell Biol**, 2003. 4(9): p. 700-11.

12. Mustelin, T., T. Vang, and N. Bottini, *Protein tyrosine phosphatases and the immune response.* **Nat Rev Immunol**, 2005. 5(1): p. 43-57.

13. Tonks, N.K., *Protein tyrosine phosphatases: from genes, to function, to disease.* **Nat Rev Mol Cell Biol**, 2006. 7(11): p. 833-46.

14. Pao, L.I., K. Badour, K.A. Siminovitch, and B.G. Neel, *Nonreceptor protein-tyrosine phosphatases in immune cell signaling.* **Annu Rev Immunol**, 2007. 25: p. 473-523.

15. Halle, M., M.L. Tremblay, and T.C. Meng, *Protein tyrosine phosphatases: emerging regulators of apoptosis.* **Cell Cycle**, 2007. 6(22): p. 2773-81.

16. Hunter, T., *Tyrosine phosphorylation: thirty years and counting.* **Curr Opin Cell Biol**, 2009. 21(2): p. 140-6.

17. Rhee, I. and A. Veillette, *Protein tyrosine phosphatases in lymphocyte activation and autoimmunity.* **Nat Immunol**, 2012. 13(5): p. 439-47.

18. Cohen, P. and D.R. Alessi, *Kinase drug discovery--what's next in the field?* **ACS Chem Biol**, 2013. 8(1): p. 96-104.

19. Tautz, L., M. Pellecchia, and T. Mustelin, *Targeting the PTPome in human disease.* **Expert Opin Ther Targets**, 2006. 10(1): p. 157-77.

20. Tonks, N.K., *Protein tyrosine phosphatases--from housekeeping enzymes to master regulators of signal transduction.* **FEBS J**, 2013. 280(2): p. 346-78.

21. Andersen, J.N., O.H. Mortensen, G.H. Peters, P.G. Drake, L.F. Iversen, O.H. Olsen, P.G. Jansen, H.S. Andersen, N.K. Tonks, and N.P. Moller, *Structural and evolutionary relationships among protein tyrosine phosphatase domains.* **Mol Cell Biol**, 2001. 21(21): p. 7117-36.

22. Pulido, R. and R. Hooft van Huijsduijnen, *Protein tyrosine phosphatases: dual-specificity phosphatases in health and disease.* **FEBS J,** 2008. 275(5): p. 848-66.

23. Bottini, N., E. Bottini, F. Gloria-Bottini, and T. Mustelin, *Low-molecular-weight protein tyrosine phosphatase and human disease: in search of biochemical mechanisms.* **Arch Immunol Ther Exp** (Warsz), 2002. 50(2): p. 95-104.

24. Tonks, N.K. and B.G. Neel, *Combinatorial control of the specificity of protein tyrosine phosphatases.* **Curr Opin Cell Biol**, 2001. 13(2): p. 182-95.

25. Cousin, W., A. Courseaux, A. Ladoux, C. Dani, and P. Peraldi, *Cloning of hOST-PTP: the only example of a protein-tyrosine-phosphatase the function of which has been lost between rodent and human.* **Biochem Biophys Res Commun**, 2004. 321(1): p. 259-65.

26. MacKay, C., A.C. Declais, C. Lundin, A. Agostinho, A.J. Deans, T.J. MacArtney, K. Hofmann, A. Gartner, S.C. West, T. Helleday, D.M. Lilley, and J. Rouse,

*Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2.* **Cell**, 2010. 142(1): p. 65-76.

27. Zhang, Z.Y., Y. Wang, and J.E. Dixon, *Dissecting the catalytic mechanism of protein-tyrosine phosphatases.* **Proc Natl Acad Sci U S A**, 1994. 91(5): p. 1624-7.

28. Zhang, Z.Y., *Kinetic and mechanistic characterization of a mammalian protein-tyrosine phosphatase, PTP1.* **J Biol Chem**, 1995. 270(19): p. 11199-204.

29. Denu, J.M., J.A. Stuckey, M.A. Saper, and J.E. Dixon, *Form and function in protein dephosphorylation.* **Cell**, 1996. 87(3): p. 361-4.

30. Denu, J.M., D.L. Lohse, J. Vijayalakshmi, M.A. Saper, and J.E. Dixon, *Visualization of intermediate and transition-state structures in protein-tyrosine phosphatase catalysis.* **Proc Natl Acad Sci U S A**, 1996. 93(6): p. 2493-8.

31. Stuckey, J.A., H.L. Schubert, E.B. Fauman, Z.Y. Zhang, J.E. Dixon, and M.A. Saper, *Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 A and the complex with tungstate.* **Nature**, 1994. 370(6490): p. 571-5.

32. Hengge, A.C., J.M. Denu, and J.E. Dixon, *Transition-state structures for the native dual-specific phosphatase VHR and D92N and S131A mutants. Contributions to the driving force for catalysis.* **Biochemistry**, 1996. 35(22): p. 7084-92.

33. Barford, D., A.J. Flint, and N.K. Tonks, *Crystal structure of human protein tyrosine phosphatase 1B.* **Science**, 1994. 263(5152): p. 1397-404.

34. Jia, Z., D. Barford, A.J. Flint, and N.K. Tonks, *Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B.* **Science**, 1995. 268(5218): p. 1754-8.

35. Zhang, Z.Y., *Protein tyrosine phosphatases: structure and function, substrate specificity, and inhibitor development.* **Annu Rev Pharmacol Toxicol**, 2002. 42: p. 209-34.

36. Zhang, Z.Y., D. Maclean, D.J. McNamara, T.K. Sawyer, and J.E. Dixon, *Protein tyrosine phosphatase substrate specificity: size and phosphotyrosine positioning requirements in peptide substrates.* **Biochemistry**, 1994. **33**(8): p. 2285-90.

37. Mustelin, T., L. Tautz, and R. Page, *Structure of the hematopoietic tyrosine phosphatase (HePTP) catalytic domain: structure of a KIM phosphatase with phosphate bound at the active site.* **J Mol Biol**, 2005. 354(1): p. 150-63.

38. Barr, A.J., E. Ugochukwu, W.H. Lee, O.N. King, P. Filippakopoulos, I. Alfano, P. Savitsky, N.A. Burgess-Brown, S. Muller, and S. Knapp, *Large-scale structural analysis of the classical human protein tyrosine phosphatome.* **Cell**, 2009. 136(2): p. 352-63.

39. Wiesmann, C., K.J. Barr, J. Kung, J. Zhu, D.A. Erlanson, W. Shen, B.J. Fahr, M. Zhong, L. Taylor, M. Randal, R.S. McDowell, and S.K. Hansen, *Allosteric inhibition of protein tyrosine phosphatase 1B.* **Nat Struct Mol Biol**, 2004. 11(8): p. 730-7.

40. Kamerlin, S.C., R. Rucker, and S. Boresch, *A molecular dynamics study of WPD-loop flexibility in PTP1B.* **Biochem Biophys Res Commun**, 2007. 356(4): p. 1011-6.

41. Tautz, L., D.A. Critton, and S. Grotegut, *Protein tyrosine phosphatases: structure, function, and implication in human disease.* **Methods Mol Biol**, 2013. 1053: p. 179-221.

42. Zhang, Z.Y. and J.E. Dixon, *Active site labeling of the Yersinia protein tyrosine phosphatase: the determination of the pKa of the active site cysteine and the function of the conserved histidine 402.* **Biochemistry**, 1993. 32(36): p. 9340-5.

43. Tonks, N.K., *Redox redux: revisiting PTPs and the control of cell signaling.* **Cell**, 2005. 121(5): p. 667-70.

44. Chen, Y.Y., H.M. Chu, K.T. Pan, C.H. Teng, D.L. Wang, A.H. Wang, K.H. Khoo, and T.C. Meng, *Cysteine S-nitrosylation protects protein-tyrosine phosphatase 1B against oxidation-induced permanent inactivation.* **J Biol Chem**, 2008. 283(50): p. 35265-72.

45. Krishnan, N., C. Fu, D.J. Pappin, and N.K. Tonks, *H2S-Induced sulfhydration of the phosphatase PTP1B and its role in the endoplasmic reticulum stress response.* **Sci Signal**, 2011. 4(203): p. ra86.

46. Keng, Y.F., L. Wu, and Z.Y. Zhang, *Probing the function of the conserved tryptophan in the flexible loop of the Yersinia protein-tyrosine phosphatase.* **Eur J Biochem**, 1999. 259(3): p. 809-14.

47. Hoff, R.H., A.C. Hengge, L. Wu, Y.F. Keng, and Z.Y. Zhang, *Effects on general acid catalysis from mutations of the invariant tryptophan and arginine residues in the protein tyrosine phosphatase from Yersinia.* **Biochemistry**, 2000. 39(1): p. 46-54.

48. Kurkcuoglu, Z., A. Bakan, D. Kocaman, I. Bahar, and P. Doruker, *Coupling between catalytic loop motions and enzyme global dynamics.* **PLoS Comput Biol**, 2012. 8(9): p. e1002705.

49. Cui, L., W.P. Yu, H.J. DeAizpurua, R.S. Schmidli, and C.J. Pallen, *Cloning and characterization of islet cell antigen-related protein-tyrosine phosphatase (PTP), a novel receptor-like PTP and autoantigen in insulin-dependent diabetes.* **J Biol Chem**, 1996. 271(40): p. 24817-23.

50. Eswaran, J., J.P. von Kries, B. Marsden, E. Longman, J.E. Debreczeni, E. Ugochukwu, A. Turnbull, W.H. Lee, S. Knapp, and A.J. Barr, *Crystal structures and inhibitor identification for PTPN5, PTPRR and PTPN7: a family of human MAPK-specific protein tyrosine phosphatases.* **Biochem J**, 2006. 395(3): p. 483-91.

51. Iversen, L.F., K.B. Moller, A.K. Pedersen, G.H. Peters, A.S. Petersen, H.S. Andersen, S. Branner, S.B. Mortensen, and N.P. Moller, *Structure determination of T cell protein-tyrosine phosphatase.* **J Biol Chem**, 2002. 277(22): p. 19982-90.

52. Yang, J., Z. Cheng, T. Niu, X. Liang, Z.J. Zhao, and G.W. Zhou, *Structural basis for substrate specificity of protein-tyrosine phosphatase SHP-1.* **J Biol Chem**, 2000. 275(6): p. 4066-71.

53. Asante-Appiah, E., S. Patel, C. Desponts, J.M. Taylor, C. Lau, C. Dufresne, M. Therien, R. Friesen, J.W. Becker, Y. Leblanc, B.P. Kennedy, and G. Scapin, *Conformation-assisted inhibition of protein-tyrosine phosphatase-1B elicits inhibitor selectivity over T-cell protein-tyrosine phosphatase.* **J Biol Chem**, 2006. 281(12): p. 8010-5.

54. Critton, D.A., L. Tautz, and R. Page, *Visualizing active-site dynamics in single crystals of HePTP: opening of the WPD loop involves coordinated movement of the E loop.* **J Mol Biol**, 2011. 405(3): p. 619-29.

55. Xie, L., Y.L. Zhang, and Z.Y. Zhang, *Design and characterization of an improved protein tyrosine phosphatase substrate-trapping mutant.* **Biochemistry**, 2002. 41(12): p. 4032-9.

56. Zhao, Y., L. Wu, S.J. Noh, K.L. Guan, and Z.Y. Zhang, *Altering the nucleophile specificity of a protein-tyrosine phosphatase-catalyzed reaction. Probing the function of the invariant glutamine residues.* **J Biol Chem**, 1998. 273(10): p. 5484-92.

57. Pedersen, A.K., X.L. Guo, K.B. Moller, G.H. Peters, H.S. Andersen, J.S. Kastrup, S.B. Mortensen, L.F. Iversen, Z.Y. Zhang, and N.P. Moller, *Residue 182 influences*

*the second step of protein-tyrosine phosphatase-mediated catalysis.* **Biochem J**, 2004. 378(Pt 2): p. 421-33.

58. Feng, G.S., C.C. Hui, and T. Pawson, *SH2-containing phosphotyrosine phosphatase as a target of protein-tyrosine kinases.* **Science**, 1993. 259(5101): p. 1607-11.

59. Freeman, R.M., Jr., J. Plutzky, and B.G. Neel, *Identification of a human src homology 2-containing protein-tyrosine-phosphatase: a putative homolog of Drosophila corkscrew.* **Proc Natl Acad Sci U S A**, 1992. 89(23): p. 11239-43.

60. Ahmad, S., D. Banville, Z. Zhao, E.H. Fischer, and S.H. Shen, *A widely expressed human protein-tyrosine phosphatase containing src homology 2 domains.* **Proc Natl Acad Sci U S A**, 1993. 90(6): p. 2197-201.

61. Vogel, W., R. Lammers, J. Huang, and A. Ullrich, *Activation of a phosphotyrosine phosphatase by tyrosine phosphorylation.* **Science**, 1993. 259(5101): p. 1611-4.

62. Adachi, M., M. Sekiya, T. Miyachi, K. Matsuno, Y. Hinoda, K. Imai, and A. Yachi, *Molecular cloning of a novel protein-tyrosine phosphatase SH-PTP3 with sequence similarity to the src-homology region 2.* **FEBS Lett**, 1992. 314(3): p. 335-9.

63. Mohi, M.G. and B.G. Neel, *The role of Shp2 (PTPN11) in cancer.* **Curr Opin Genet Dev**, 2007. 17(1): p. 23-30.

64. Neel, B.G., H. Gu, and L. Pao, *The 'Shp'ing news: SH2 domain-containing tyrosine phosphatases in cell signaling.* **Trends Biochem Sci**, 2003. 28(6): p. 284-93.

65. Tartaglia, M., E.L. Mehler, R. Goldberg, G. Zampino, H.G. Brunner, H. Kremer, I. van der Burgt, A.H. Crosby, A. Ion, S. Jeffery, K. Kalidas, M.A. Patton, R.S. Kucherlapati, and B.D. Gelb, *Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome.* **Nat Genet**, 2001. 29(4): p. 465-8.

66. Noonan, J.A., *Hypertelorism with Turner phenotype. A new syndrome with associated congenital heart disease.* **Am J Dis Child**, 1968. 116(4): p. 373-80.

67. Allanson, J.E., *Noonan syndrome.* **J Med Genet**, 1987. 24(1): p. 9-13.

68. Marino, B., M.C. Digilio, A. Toscano, A. Giannotti, and B. Dallapiccola, *Congenital heart diseases in children with Noonan syndrome: An expanded cardiac spectrum with high prevalence of atrioventricular canal.* **J Pediatr**, 1999. 135(6): p. 703-6.

69. Tartaglia, M. and B.D. Gelb, *Noonan syndrome and related disorders: genetics and pathogenesis.* **Annu Rev Genomics Hum Genet**, 2005. 6: p. 45-68.

70. Opitz, J.M., *The Noonan syndrome.* **Am J Med Genet**, 1985. 21(3): p. 515-8.

71. Tartaglia, M., C.M. Niemeyer, A. Fragale, X. Song, J. Buechner, A. Jung, K. Hahlen, H. Hasle, J.D. Licht, and B.D. Gelb, *Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia.* **Nat Genet**, 2003. 34(2): p. 148-50.

72. Bentires-Alj, M., J.G. Paez, F.S. David, H. Keilhack, B. Halmos, K. Naoki, J.M. Maris, A. Richardson, A. Bardelli, D.J. Sugarbaker, W.G. Richards, J. Du, L. Girard, J.D. Minna, M.L. Loh, D.E. Fisher, V.E. Velculescu, B. Vogelstein, M. Meyerson, W.R. Sellers, and B.G. Neel, *Activating mutations of the noonan syndrome-associated SHP2/PTPN11 gene in human solid tumors and adult acute myelogenous leukemia.* **Cancer Res**, 2004. 64(24): p. 8816-20.

73. Grossmann, K.S., M. Rosario, C. Birchmeier, and W. Birchmeier, *The tyrosine phosphatase Shp2 in development and cancer.* **Adv Cancer Res**, 2010. 106: p. 53-89.

74. Chan, R.J. and G.S. Feng, *PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase.* **Blood**, 2007. 109(3): p. 862-7.

75. Chan, G., D. Kalaitzidis, and B.G. Neel, *The tyrosine phosphatase Shp2 (PTPN11) in cancer.* **Cancer Metastasis Rev**, 2008. 27(2): p. 179-92.

76. Deb, T.B., L. Wong, D.S. Salomon, G. Zhou, J.E. Dixon, J.S. Gutkind, S.A. Thompson, and G.R. Johnson, *A common requirement for the catalytic activity and both SH2 domains of SHP-2 in mitogen-activated protein (MAP) kinase activation by the ErbB family of receptors. A specific role for SHP-2 in map, but not c-Jun amino-terminal kinase activation.* **J Biol Chem**, 1998. 273(27): p. 16643-6.

77. Van Vactor, D., A.M. O'Reilly, and B.G. Neel, *Genetic analysis of protein tyrosine phosphatases.* **Curr Opin Genet Dev**, 1998. 8(1): p. 112-26.

78. Araki, T., H. Nawa, and B.G. Neel, *Tyrosyl phosphorylation of Shp2 is required for normal ERK activation in response to some, but not all, growth factors.* **J Biol Chem**, 2003. 278(43): p. 41677-84.

79. Walter, A.O., Z.Y. Peng, and C.A. Cartwright, *The Shp-2 tyrosine phosphatase activates the Src tyrosine kinase by a non-enzymatic mechanism.* **Oncogene**, 1999. 18(11): p. 1911-20.

80. Hof, P., S. Pluskey, S. Dhe-Paganon, M.J. Eck, and S.E. Shoelson, *Crystal structure of the tyrosine phosphatase SHP-2.* **Cell**, 1998. 92(4): p. 441-50.

81. Barford, D. and B.G. Neel, *Revealing mechanisms for SH2 domain mediated regulation of the protein tyrosine phosphatase SHP-2.* **Structure**, 1998. 6(3): p. 249-54.

82. O'Reilly, A.M., S. Pluskey, S.E. Shoelson, and B.G. Neel, *Activated mutants of SHP-2 preferentially induce elongation of Xenopus animal caps.* **Mol Cell Biol**, 2000. 20(1): p. 299-311.

83. Matozaki, T., Y. Murata, Y. Saito, H. Okazawa, and H. Ohnishi, *Protein tyrosine phosphatase SHP-2: a proto-oncogene product that promotes Ras activation.* **Cancer Sci**, 2009. 100(10): p. 1786-93.

84. Tiganis, T. and A.M. Bennett, *Protein tyrosine phosphatase function: the substrate perspective.* **Biochem J**, 2007. 402(1): p. 1-15.

85. Bard-Chapeau, E.A., J. Yuan, N. Droin, S. Long, E.E. Zhang, T.V. Nguyen, and G.S. Feng, *Concerted functions of Gab1 and Shp2 in liver regeneration and hepatoprotection.* **Mol Cell Biol**, 2006. 26(12): p. 4664-74.

86. Agazie, Y.M. and M.J. Hayman, *Molecular mechanism for a role of SHP2 in epidermal growth factor receptor signaling.* **Mol Cell Biol**, 2003. 23(21): p. 7875-86.

87. Agazie, Y.M. and M.J. Hayman, *Development of an efficient "substrate-trapping" mutant of Src homology phosphotyrosine phosphatase 2 and identification of the epidermal growth factor receptor, Gab1, and three other proteins as target substrates.* **J Biol Chem**, 2003. 278(16): p. 13952-8.

88. Montagner, A., A. Yart, M. Dance, B. Perret, J.P. Salles, and P. Raynal, *A novel role for Gab1 and SHP2 in epidermal growth factor-induced Ras activation.* **J Biol Chem**, 2005. 280(7): p. 5350-60.

89. Zhang, S.Q., W. Yang, M.I. Kontaridis, T.G. Bivona, G. Wen, T. Araki, J. Luo, J.A. Thompson, B.L. Schraven, M.R. Philips, and B.G. Neel, *Shp2 regulates SRC family kinase activity and Ras/Erk activation by controlling Csk recruitment.* **Mol Cell**, 2004. 13(3): p. 341-55.

90. Klinghoffer, R.A. and A. Kazlauskas, *Identification of a putative Syp substrate, the PDGF beta receptor.* **J Biol Chem**, 1995. 270(38): p. 22208-17.

91. Cleghon, V., P. Feldmann, C. Ghiglione, T.D. Copeland, N. Perrimon, D.A. Hughes, and D.K. Morrison, *Opposing actions of CSW and RasGAP modulate the strength of Torso RTK signaling in the Drosophila terminal pathway.* **Mol Cell**, 1998. 2(6): p. 719-27.

92. Hanafusa, H., S. Torii, T. Yasunaga, K. Matsumoto, and E. Nishida, *Shp2, an SH2-containing protein-tyrosine phosphatase, positively regulates receptor tyrosine kinase signaling by dephosphorylating and inactivating the inhibitor Sprouty.* **J Biol Chem**, 2004. 279(22): p. 22992-5.

93. Jarvis, L.A., S.J. Toering, M.A. Simon, M.A. Krasnow, and R.K. Smith-Bolton, *Sprouty proteins are in vivo targets of Corkscrew/SHP-2 tyrosine phosphatases.* **Development**, 2006. 133(6): p. 1133-42.

94. Bunda, S., P. Heir, T. Srikumar, J.D. Cook, K. Burrell, Y. Kano, J.E. Lee, G. Zadeh, B. Raught, and M. Ohh, *Src promotes GTPase activity of Ras via tyrosine 32 phosphorylation.* **Proc Natl Acad Sci U S A**, 2014. 111(36): p. E3785-94.

95. Bunda, S., K. Burrell, P. Heir, L. Zeng, A. Alamsahebpour, Y. Kano, B. Raught, Z.Y. Zhang, G. Zadeh, and M. Ohh, *Inhibition of SHP2-mediated dephosphorylation of Ras suppresses oncogenesis.* **Nat Commun**, 2015. 6: p. 8859.

96. Zhang, S.Q., W.G. Tsiaras, T. Araki, G. Wen, L. Minichiello, R. Klein, and B.G. Neel, *Receptor-specific regulation of phosphatidylinositol 3'-kinase activation by the protein tyrosine phosphatase Shp2.* **Mol Cell Biol**, 2002. 22(12): p. 4062-72.

97. Mattoon, D.R., B. Lamothe, I. Lax, and J. Schlessinger, *The docking protein Gab1 is the primary mediator of EGF-stimulated activation of the PI-3K/Akt cell survival pathway.* **BMC Biol**, 2004. 2: p. 24.

98. Shi, Z.Q., W. Lu, and G.S. Feng, *The Shp-2 tyrosine phosphatase has opposite effects in mediating the activation of extracellular signal-regulated and c-Jun NH2-terminal mitogen-activated protein kinases.* **J Biol Chem**, 1998. 273(9): p. 4904-8.

99. You, M., L.M. Flick, D. Yu, and G.S. Feng, *Modulation of the nuclear factor kappa B pathway by Shp-2 tyrosine phosphatase in mediating the induction of interleukin (IL)-6 by IL-1 or tumor necrosis factor.* **J Exp Med**, 2001. 193(1): p. 101-10.

100. Schoenwaelder, S.M., L.A. Petch, D. Williamson, R. Shen, G.S. Feng, and K. Burridge, *The protein tyrosine phosphatase Shp-2 regulates RhoA activity.* **Curr Biol**, 2000. 10(23): p. 1523-6.

101. Uhlen, P., P.M. Burch, C.I. Zito, M. Estrada, B.E. Ehrlich, and A.M. Bennett, *Gain-of-function/Noonan syndrome SHP-2/Ptpn11 mutants enhance calcium oscillations and impair NFAT signaling.* **Proc Natl Acad Sci U S A**, 2006. 103(7): p. 2160-5.

102. Fruhbeck, G., *Intracellular signalling pathways activated by leptin.* **Biochem J**, 2006. 393(Pt 1): p. 7-20.

103. Schubbert, S., G. Bollag, N. Lyubynska, H. Nguyen, C.P. Kratz, M. Zenker, C.M. Niemeyer, A. Molven, and K. Shannon, *Biochemical and functional characterization of germ line KRAS mutations.* **Mol Cell Biol**, 2007. 27(22): p. 7765-70.

104. Sarkozy, A., M.C. Digilio, and B. Dallapiccola, *Leopard syndrome.* **Orphanet J Rare Dis**, 2008. 3: p. 13.

105. Passmore, S.J., J.M. Chessells, H. Kempski, I.M. Hann, P.A. Brownbill, and C.A. Stiller, *Paediatric myelodysplastic syndromes and juvenile myelomonocytic leukaemia in the UK: a population-based study of incidence and survival.* **Br J Haematol**, 2003. 121(5): p. 758-67.

106. Keilhack, H., F.S. David, M. McGregor, L.C. Cantley, and B.G. Neel, *Diverse biochemical properties of Shp2 mutants. Implications for disease phenotypes.* **J Biol Chem**, 2005. 280(35): p. 30984-93.

107. Schubbert, S., K. Lieuw, S.L. Rowe, C.M. Lee, X. Li, M.L. Loh, D.W. Clapp, and K.M. Shannon, *Functional analysis of leukemia-associated PTPN11 mutations in primary hematopoietic cells.* **Blood**, 2005. 106(1): p. 311-7.

108. Kontaridis, M.I., K.D. Swanson, F.S. David, D. Barford, and B.G. Neel, *PTPN11 (Shp2) mutations in LEOPARD syndrome have dominant negative, not activating, effects.* **J Biol Chem**, 2006. 281(10): p. 6785-92.

109. Yu, Z.H., J. Xu, C.D. Walls, L. Chen, S. Zhang, R. Zhang, L. Wu, L. Wang, S. Liu, and Z.Y. Zhang, *Structural and mechanistic insights into LEOPARD syndrome-associated SHP2 mutations.* **J Biol Chem**, 2013. 288(15): p. 10472-82.

110. Qiu, W., X. Wang, V. Romanov, A. Hutchinson, A. Lin, M. Ruzanov, K.P. Battaile, E.F. Pai, B.G. Neel, and N.Y. Chirgadze, *Structural insights into Noonan/LEOPARD syndrome-related mutants of protein-tyrosine phosphatase SHP2 (PTPN11).* **BMC Struct Biol**, 2014. 14: p. 10.

111. Kitayama, H., Y. Kanakura, T. Furitsu, T. Tsujimura, K. Oritani, H. Ikeda, H. Sugahara, H. Mitsui, Y. Kanayama, Y. Kitamura, and et al., *Constitutively activating mutations of c-kit receptor tyrosine kinase confer factor-independent*

*growth and tumorigenicity of factor-dependent hematopoietic cell lines.* **Blood**, 1995. 85(3): p. 790-8.

112. Munugalavadla, V., E.C. Sims, J. Borneo, R.J. Chan, and R. Kapur, *Genetic and pharmacologic evidence implicating the p85 alpha, but not p85 beta, regulatory subunit of PI3K and Rac2 GTPase in regulating oncogenic KIT-induced transformation in acute myeloid leukemia and systemic mastocytosis.* **Blood**, 2007. 110(5): p. 1612-20.

113. Mali, R.S., P. Ma, L.F. Zeng, H. Martin, B. Ramdas, Y. He, E. Sims, S. Nabinger, J. Ghosh, N. Sharma, V. Munugalavadla, A. Chatterjee, S. Li, G. Sandusky, A.W. Craig, K.D. Bunting, G.S. Feng, R.J. Chan, Z.Y. Zhang, and R. Kapur, *Role of SHP2 phosphatase in KIT-induced transformation: identification of SHP2 as a druggable target in diseases involving oncogenic KIT.* **Blood**, 2012. 120(13): p. 2669-78.

114. Mali, R.S., R. Chan, and R. Kapur, *Targeting SHP2 phosphatase in myeloproliferative neoplasms.* **Oncotarget**, 2012. 3(10): p. 1049-51.

115. Eccles, S.A., E.O. Aboagye, S. Ali, A.S. Anderson, J. Armes, F. Berditchevski, J.P. Blaydes, K. Brennan, N.J. Brown, H.E. Bryant, N.J. Bundred, J.M. Burchell, A.M. Campbell, J.S. Carroll, R.B. Clarke, C.E. Coles, G.J. Cook, A. Cox, N.J. Curtin, L.V. Dekker, S. Silva Idos, S.W. Duffy, D.F. Easton, D.M. Eccles, D.R. Edwards, J. Edwards, D. Evans, D.F. Fenlon, J.M. Flanagan, C. Foster, W.M. Gallagher, M. Garcia-Closas, J.M. Gee, A.J. Gescher, V. Goh, A.M. Groves, A.J. Harvey, M. Harvie, B.T. Hennessy, S. Hiscox, I. Holen, S.J. Howell, A. Howell, G. Hubbard, N. Hulbert-Williams, M.S. Hunter, B. Jasani, L.J. Jones, T.J. Key, C.C. Kirwan, A. Kong, I.H. Kunkler, S.P. Langdon, M.O. Leach, D.J. Mann, J.F. Marshall, L. Martin, S.G. Martin, J.E. Macdougall, D.W. Miles, W.R. Miller, J.R. Morris, S.M. Moss, P. Mullan, R. Natrajan, J.P. O'Connor, R. O'Connor, C. Palmieri, P.D. Pharoah, E.A. Rakha, E. Reed, S.P. Robinson, E. Sahai, J.M. Saxton, P. Schmid, M.J. Smalley, V. Speirs, R. Stein, J. Stingl, C.H. Streuli, A.N. Tutt, G. Velikova, R.A. Walker, C.J. Watson, K.J. Williams, L.S. Young, and A.M. Thompson, *Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer.* **Breast Cancer Res**, 2013. 15(5): p. R92.

116. Pereira, B., S.F. Chin, O.M. Rueda, H.K. Vollan, E. Provenzano, H.A. Bardwell, M. Pugh, L. Jones, R. Russell, S.J. Sammut, D.W. Tsui, B. Liu, S.J. Dawson, J. Abraham, H. Northen, J.F. Peden, A. Mukherjee, G. Turashvili, A.R. Green, S. McKinney, A. Oloumi, S. Shah, N. Rosenfeld, L. Murphy, D.R. Bentley, I.O. Ellis, A. Purushotham, S.E. Pinder, A.L. Borresen-Dale, H.M. Earl, P.D. Pharoah, M.T. Ross, S. Aparicio, and C. Caldas, *The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes.* **Nat Commun**, 2016. 7: p. 11479.

117. Bentires-Alj, M., S.G. Gil, R. Chan, Z.C. Wang, Y. Wang, N. Imanaka, L.N. Harris, A. Richardson, B.G. Neel, and H. Gu, *A role for the scaffolding adapter GAB2 in breast cancer.* **Nat Med**, 2006. 12(1): p. 114-21.

118. Bocanegra, M., A. Bergamaschi, Y.H. Kim, M.A. Miller, A.B. Rajput, J. Kao, A. Langerod, W. Han, D.Y. Noh, S.S. Jeffrey, D.G. Huntsman, A.L. Borresen-Dale, and J.R. Pollack, *Focal amplification and oncogene dependency of GAB2 in breast cancer*. **Oncogene**, 2010. 29(5): p. 774-9.

119. Zhou, X., J. Coad, B. Ducatman, and Y.M. Agazie, *SHP2 is up-regulated in breast cancer cells and in infiltrating ductal carcinoma of the breast, implying its involvement in breast oncogenesis.* **Histopathology**, 2008. 53(4): p. 389-402.

120. Aceto, N., N. Sausgruber, H. Brinkhaus, D. Gaidatzis, G. Martiny-Baron, G. Mazzarol, S. Confalonieri, M. Quarto, G. Hu, P.J. Balwierz, M. Pachkov, S.J. Elledge, E. van Nimwegen, M.B. Stadler, and M. Bentires-Alj, *Tyrosine phosphatase SHP2 promotes breast cancer progression and maintains tumor-initiating cells via activation of key transcription factors and a positive feedback signaling loop.* **Nat Med**, 2012. 18(4): p. 529-37.

121. Aceto, N. and M. Bentires-Alj, *Targeting protein-tyrosine phosphatases in breast cancer*. **Oncotarget**, 2012. 3(5): p. 514-5.

122. Lacey, J.V., Jr., A.R. Kreimer, S.S. Buys, P.M. Marcus, S.C. Chang, M.F. Leitzmann, R.N. Hoover, P.C. Prorok, C.D. Berg, P. Hartge, L.C. Prostate, and T. Ovarian Cancer Screening Trial Project, *Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort.* **BMC Cancer**, 2009. 9: p. 84.

123. Hatakeyama, M. and H. Higashi, *Helicobacter pylori CagA: a new paradigm for bacterial carcinogenesis.* **Cancer Sci**, 2005. 96(12): p. 835-43.

124. Hatakeyama, M., *Oncogenic mechanisms of the Helicobacter pylori CagA protein.* **Nat Rev Cancer**, 2004. 4(9): p. 688-94.

125. Druker, B.J., S. Tamura, E. Buchdunger, S. Ohno, G.M. Segal, S. Fanning, J. Zimmermann, and N.B. Lydon, *Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells.* **Nat Med**, 1996. 2(5): p. 561-6.

126. Hynes, N.E. and H.A. Lane, *ERBB receptors and cancer: the complexity of targeted inhibitors.* **Nat Rev Cancer**, 2005. 5(5): p. 341-54.

127. Engelman, J.A. and J. Settleman, *Acquired resistance to tyrosine kinase inhibitors during cancer therapy*. **Curr Opin Genet Dev**, 2008. 18(1): p. 73-9.

128. Rexer, B.N. and C.L. Arteaga, *Intrinsic and acquired resistance to HER2-targeted therapies in HER2 gene-amplified breast cancer: mechanisms and clinical implications.* **Crit Rev Oncog**, 2012. 17(1): p. 1-16.

129. Gharwan, H. and H. Groninger, *Kinase inhibitors and monoclonal antibodies in oncology: clinical implications.* **Nat Rev Clin Oncol**, 2016. 13(4): p. 209-27.

130. Lazo, J.S. and E.R. Sharlow, *Drugging Undruggable Molecular Cancer Targets.* **Annu Rev Pharmacol Toxicol**, 2016. 56: p. 23-40.

131. Andersen, J.N., P.G. Jansen, S.M. Echwald, O.H. Mortensen, T. Fukada, R. Del Vecchio, N.K. Tonks, and N.P. Moller, *A genomic perspective on protein tyrosine phosphatases: gene structure, pseudogenes, and genetic disease linkage.* **FASEB J**, 2004. 18(1): p. 8-30.

132. Arrowsmith, C.H., J.E. Audia, C. Austin, J. Baell, J. Bennett, J. Blagg, C. Bountra, P.E. Brennan, P.J. Brown, M.E. Bunnage, C. Buser-Doepner, R.M. Campbell, A.J. Carter, P. Cohen, R.A. Copeland, B. Cravatt, J.L. Dahlin, D. Dhanak, A.M. Edwards, M. Frederiksen, S.V. Frye, N. Gray, C.E. Grimshaw, D. Hepworth, T. Howe, K.V. Huber, J. Jin, S. Knapp, J.D. Kotz, R.G. Kruger, D. Lowe, M.M. Mader, B. Marsden, A. Mueller-Fahrnow, S. Muller, R.C. O'Hagan, J.P. Overington, D.R. Owen, S.H. Rosenberg, B. Roth, R. Ross, M. Schapira, S.L. Schreiber, B. Shoichet, M. Sundstrom, G. Superti-Furga, J. Taunton, L. Toledo-Sherman, C. Walpole, M.A. Walters, T.M. Willson, P. Workman, R.N. Young, and W.J. Zuercher, *The promise and peril of chemical probes.* **Nat Chem Biol**, 2015. 11(8): p. 536-41.

133. Frezard, F., C. Demicheli, and R.R. Ribeiro, *Pentavalent antimonials: new perspectives for old drugs.* **Molecules**, 2009. 14(7): p. 2317-36.

134. Pathak, M.K. and T. Yi, *Sodium stibogluconate is a potent inhibitor of protein tyrosine phosphatases and augments cytokine responses in hemopoietic cell lines.* **J Immunol**, 2001. 167(6): p. 3391-7.

135. Yi, T., P. Elson, M. Mitsuhashi, B. Jacobs, E. Hollovary, T.G. Budd, T. Spiro, P. Triozzi, and E.C. Borden, *Phosphatase inhibitor, sodium stibogluconate, in combination with interferon (IFN) alpha 2b: phase I trials to identify pharmacodynamic and clinical effects.* **Oncotarget**, 2011. 2(12): p. 1155-64.

136. Chen, L., S.S. Sung, M.L. Yip, H.R. Lawrence, Y. Ren, W.C. Guida, S.M. Sebti, N.J. Lawrence, and J. Wu, *Discovery of a novel shp2 protein tyrosine phosphatase inhibitor.* **Mol Pharmacol**, 2006. 70(2): p. 562-70.

137. Luo, Q., Y. Sun, F.Y. Gong, W. Liu, W. Zheng, Y. Shen, Z.C. Hua, and Q. Xu, *Blocking initial infiltration of pioneer CD8(+) T-cells into the CNS via inhibition*

*of SHP-2 ameliorates experimental autoimmune encephalomyelitis in mice.* **Br J Pharmacol**, 2014. 171(7): p. 1706-21.

138. Lawrence, H.R., R. Pireddu, L. Chen, Y. Luo, S.S. Sung, A.M. Szymanski, M.L. Yip, W.C. Guida, S.M. Sebti, J. Wu, and N.J. Lawrence, *Inhibitors of Src homology-2 domain containing protein tyrosine phosphatase-2 (Shp2) based on oxindole scaffolds.* **J Med Chem**, 2008. 51(16): p. 4948-56.

139. Noren-Muller, A., I. Reis-Correa, Jr., H. Prinz, C. Rosenbaum, K. Saxena, H.J. Schwalbe, D. Vestweber, G. Cagna, S. Schunk, O. Schwarz, H. Schiewe, and H. Waldmann, *Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis.* **Proc Natl Acad Sci U S A**, 2006. 103(28): p. 10606-11.

140. Hellmuth, K., S. Grosskopf, C.T. Lum, M. Wurtele, N. Roder, J.P. von Kries, M. Rosario, J. Rademann, and W. Birchmeier, *Specific inhibitors of the protein tyrosine phosphatase Shp2 identified by high-throughput docking.* **Proc Natl Acad Sci U S A**, 2008. 105(20): p. 7275-80.

141. Schramm, C., M.A. Edwards, and M. Krenz, *New approaches to prevent LEOPARD syndrome-associated cardiac hypertrophy by specifically targeting Shp2-dependent signaling.* **J Biol Chem**, 2013. 288(25): p. 18335-44.

142. Xu, J., L.F. Zeng, W. Shen, J.J. Turchi, and Z.Y. Zhang, *Targeting SHP2 for EGFR inhibitor resistant non-small cell lung carcinoma.* **Biochem Biophys Res Commun**, 2013. 439(4): p. 586-90.

143. Zhang, X., Y. He, S. Liu, Z. Yu, Z.X. Jiang, Z. Yang, Y. Dong, S.C. Nabinger, L. Wu, A.M. Gunawan, L. Wang, R.J. Chan, and Z.Y. Zhang, *Salicylic acid based small molecule inhibitor for the oncogenic Src homology-2 domain containing protein tyrosine phosphatase-2 (SHP2).* **J Med Chem**, 2010. 53(6): p. 2482-93.

144. Ravery, V., K. Fizazi, S. Oudard, L. Drouet, J.C. Eymard, S. Culine, G. Gravis, C. Hennequin, and M. Zerbib, *The use of estramustine phosphate in the modern management of advanced prostate cancer.* **BJU Int**, 2011. 108(11): p. 1782-6.

145. Krishnan, N., D. Koveal, D.H. Miller, B. Xue, S.D. Akshinthala, J. Kragelj, M.R. Jensen, C.M. Gauss, R. Page, M. Blackledge, S.K. Muthuswamy, W. Peti, and N.K. Tonks, *Targeting the disordered C terminus of PTP1B with an allosteric inhibitor.* **Nat Chem Biol**, 2014. 10(7): p. 558-66.

146. Liu, W., B. Yu, G. Xu, W.R. Xu, M.L. Loh, L.D. Tang, and C.K. Qu, *Identification of cryptotanshinone as an inhibitor of oncogenic protein tyrosine phosphatase SHP2 (PTPN11).* **J Med Chem**, 2013. 56(18): p. 7212-21.

147. Wang, Q., D. Dube, R.W. Friesen, T.G. LeRiche, K.P. Bateman, L. Trimble, J. Sanghara, R. Pollex, C. Ramachandran, M.J. Gresser, and Z. Huang, *Catalytic inactivation of protein tyrosine phosphatase CD45 and protein tyrosine phosphatase 1B by polyaromatic quinones.* **Biochemistry**, 2004. 43(14): p. 4294-303.

148. Liu, S., Z. Yu, X. Yu, S.X. Huang, Y. Luo, L. Wu, W. Shen, Z. Yang, L. Wang, A.M. Gunawan, R.J. Chan, B. Shen, and Z.Y. Zhang, *SHP2 is a target of the immunosuppressant tautomycetin.* **Chem Biol**, 2011. 18(1): p. 101-10.

149. Chen, Y.N., M.J. LaMarche, H.M. Chan, P. Fekkes, J. Garcia-Fortanet, M.G. Acker, B. Antonakos, C.H. Chen, Z. Chen, V.G. Cooke, J.R. Dobson, Z. Deng, F. Fei, B. Firestone, M. Fodor, C. Fridrich, H. Gao, D. Grunenfelder, H.X. Hao, J. Jacob, S. Ho, K. Hsiao, Z.B. Kang, R. Karki, M. Kato, J. Larrow, L.R. La Bonte, F. Lenoir, G. Liu, S. Liu, D. Majumdar, M.J. Meyer, M. Palermo, L. Perez, M. Pu, E. Price, C. Quinn, S. Shakya, M.D. Shultz, J. Slisz, K. Venkatesan, P. Wang, M. Warmuth, S. Williams, G. Yang, J. Yuan, J.H. Zhang, P. Zhu, T. Ramsey, N.J. Keen, W.R. Sellers, T. Stams, and P.D. Fortin, *Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases.* **Nature**, 2016. 535(7610): p. 148-52.

150. Ran, H., R. Tsutsumi, T. Araki, and B.G. Neel, *Sticking It to Cancer with Molecular Glue for SHP2.* **Cancer Cell**, 2016. 30(2): p. 194-6.

151. Garcia Fortanet, J., C.H. Chen, Y.N. Chen, Z. Chen, Z. Deng, B. Firestone, P. Fekkes, M. Fodor, P.D. Fortin, C. Fridrich, D. Grunenfelder, S. Ho, Z.B. Kang, R. Karki, M. Kato, N. Keen, L.R. LaBonte, J. Larrow, F. Lenoir, G. Liu, S. Liu, F. Lombardo, D. Majumdar, M.J. Meyer, M. Palermo, L. Perez, M. Pu, T. Ramsey, W.R. Sellers, M.D. Shultz, T. Stams, C. Towler, P. Wang, S.L. Williams, J.H. Zhang, and M.J. LaMarche, *Allosteric Inhibition of SHP2: Identification of a Potent, Selective, and Orally Efficacious Phosphatase Inhibitor.* **J Med Chem**, 2016. 59(17): p. 7773-82.

152. McGovern, S.L., E. Caselli, N. Grigorieff, and B.K. Shoichet, *A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening.* **J Med Chem**, 2002. 45(8): p. 1712-22.

153. Denu, J.M. and K.G. Tanner, *Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulfenic acid intermediate and implications for redox regulation.* **Biochemistry**, 1998. 37(16): p. 5633-42.

154. Moretto, A.F., S.J. Kirincich, W.X. Xu, M.J. Smith, Z.K. Wan, D.P. Wilson, B.C. Follows, E. Binnun, D. Joseph-McCarthy, K. Foreman, D.V. Erbe, Y.L. Zhang, S.K. Tam, S.Y. Tam, and J. Lee, *Bicyclic and tricyclic thiophenes as protein tyrosine phosphatase 1B inhibitors.* **Bioorg Med Chem**, 2006. 14(7): p. 2162-77.

155. Thorson, M.K., D.T. Puerta, S.M. Cohen, and A.M. Barrios, *Inhibition of the lymphoid tyrosine phosphatase: the effect of zinc(II) ions and chelating ligand fragments on enzymatic activity.* **Bioorg Med Chem Lett**, 2014. 24(16): p. 4019-22.

156. Szczepankiewicz, B.G., G. Liu, P.J. Hajduk, C. Abad-Zapatero, Z. Pei, Z. Xin, T.H. Lubben, J.M. Trevillyan, M.A. Stashko, S.J. Ballaron, H. Liang, F. Huang, C.W. Hutchins, S.W. Fesik, and M.R. Jirousek, *Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy.* **J Am Chem Soc**, 2003. 125(14): p. 4087-96.

157. Liu, G., Z. Xin, Z. Pei, P.J. Hajduk, C. Abad-Zapatero, C.W. Hutchins, H. Zhao, T.H. Lubben, S.J. Ballaron, D.L. Haasch, W. Kaszubska, C.M. Rondinone, J.M. Trevillyan, and M.R. Jirousek, *Fragment screening and assembly: a highly efficient approach to a selective and cell active protein tyrosine phosphatase 1B inhibitor.* **J Med Chem**, 2003. 46(20): p. 4232-5.

158. Liu, G., B.G. Szczepankiewicz, Z. Pei, D.A. Janowick, Z. Xin, P.J. Hajduk, C. Abad-Zapatero, H. Liang, C.W. Hutchins, S.W. Fesik, S.J. Ballaron, M.A. Stashko, T. Lubben, A.K. Mika, B.A. Zinker, J.M. Trevillyan, and M.R. Jirousek, *Discovery and structure-activity relationship of oxalylarylaminobenzoic acids as inhibitors of protein tyrosine phosphatase 1B.* **J Med Chem**, 2003. 46(11): p. 2093-103.

159. Gilmartin, A.G., T.H. Faitg, M. Richter, A. Groy, M.A. Seefeld, M.G. Darcy, X. Peng, K. Federowicz, J. Yang, S.Y. Zhang, E. Minthorn, J.P. Jaworski, M. Schaber, S. Martens, D.E. McNulty, R.H. Sinnamon, H. Zhang, R.B. Kirkpatrick, N. Nevins, G. Cui, B. Pietrak, E. Diaz, A. Jones, M. Brandt, B. Schwartz, D.A. Heerding, and R. Kumar, *Allosteric Wip1 phosphatase inhibition through flap-subdomain interaction.* **Nat Chem Biol**, 2014. 10(3): p. 181-7.

160. Krueger, A.B., S.J. Dehdashti, N. Southall, J.J. Marugan, M. Ferrer, X. Li, H.L. Ford, W. Zheng, and R. Zhao, *Identification of a selective small-molecule inhibitor series targeting the eyes absent 2 (Eya2) phosphatase activity.* **J Biomol Screen**, 2013. 18(1): p. 85-96.

161. Pluskey, S., T.J. Wandless, C.T. Walsh, and S.E. Shoelson, *Potent stimulation of SH-PTP2 phosphatase activity by simultaneous occupancy of both SH2 domains.* **J Biol Chem**, 1995. 270(7): p. 2897-900.

162. Manley, P.W., P. Drueckes, G. Fendrich, P. Furet, J. Liebetanz, G. Martiny-Baron, J. Mestan, J. Trappe, M. Wartmann, and D. Fabbro, *Extended kinase profile and properties of the protein kinase inhibitor nilotinib.* **Biochim Biophys Acta**, 2010. 1804(3): p. 445-53.

163. Lund, G., S. Dudkin, D. Borkin, W. Ni, J. Grembecka, and T. Cierpicki, *Inhibition of CDC25B phosphatase through disruption of protein-protein interaction.* **ACS Chem Biol**, 2015. 10(2): p. 390-4.

164. Keseru, G.M., D.A. Erlanson, G.G. Ferenczy, M.M. Hann, C.W. Murray, and S.D. Pickett, *Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia.* **J Med Chem**, 2016.

165. Congreve, M., G. Chessari, D. Tisi, and A.J. Woodhead, *Recent developments in fragment-based drug discovery.* **J Med Chem**, 2008. 51(13): p. 3661-80.

166. Schulz, M.N. and R.E. Hubbard, *Recent progress in fragment-based lead discovery.* **Curr Opin Pharmacol**, 2009. 9(5): p. 615-21.

167. Chessari, G. and A.J. Woodhead, *From fragment to clinical candidate--a historical perspective.* **Drug Discov Today**, 2009. 14(13-14): p. 668-75.

168. Blundell, T.L., *Structure-based drug design.* **Nature**, 1996. 384(6604 Suppl): p. 23-6.

169. Spencer, R.W., *High-throughput screening of historic collections: observations on file size, biological targets, and file diversity.* **Biotechnol Bioeng**, 1998. 61(1): p. 61-7.

170. Pereira, D.A. and J.A. Williams, *Origin and evolution of high throughput screening.* **Br J Pharmacol**, 2007. 152(1): p. 53-61.

171. Pollack, S.J., K.S. Beyer, C. Lock, I. Muller, D. Sheppard, M. Lipkin, D. Hardick, P. Blurton, P.M. Leonard, P.A. Hubbard, D. Todd, C.M. Richardson, T. Ahrens, M. Baader, D.O. Hafenbradl, K. Hilyard, and R.W. Burli, *A comparative study of fragment screening methods on the p38alpha kinase: new methods, new insights.* **J Comput Aided Mol Des**, 2011. 25(7): p. 677-87.

172. Hann, M.M., A.R. Leach, and G. Harper, *Molecular complexity and its impact on the probability of finding leads for drug discovery.* **J Chem Inf Comput Sci**, 2001. 41(3): p. 856-64.

173. Hopkins, A.L., J.S. Mason, and J.P. Overington, *Can we rationally design promiscuous drugs?* **Curr Opin Struct Biol**, 2006. 16(1): p. 127-36.

174. Hajduk, P.J. and J. Greer, *A decade of fragment-based drug design: strategic advances and lessons learned.* **Nat Rev Drug Discov**, 2007. 6(3): p. 211-9.

175. Vanwetswinkel, S., R.J. Heetebrij, J. van Duynhoven, J.G. Hollander, D.V. Filippov, P.J. Hajduk, and G. Siegal, *TINS, target immobilized NMR screening: an efficient and sensitive method for ligand discovery.* **Chem Biol**, 2005. 12(2): p. 207-16.

176. Dalvit, C., P. Pevarello, M. Tato, M. Veronesi, A. Vulpetti, and M. Sundstrom, *Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water.* **J Biomol NMR**, 2000. 18(1): p. 65-8.

177. Meyer, B. and T. Peters, *NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors.* **Angew Chem Int Ed Engl**, 2003. 42(8): p. 864-90.

178. Shuker, S.B., P.J. Hajduk, R.P. Meadows, and S.W. Fesik, *Discovering high-affinity ligands for proteins: SAR by NMR.* **Science**, 1996. 274(5292): p. 1531-4.

179. Christopher, J.A., J. Brown, A.S. Dore, J.C. Errey, M. Koglin, F.H. Marshall, D.G. Myszka, R.L. Rich, C.G. Tate, B. Tehan, T. Warne, and M. Congreve, *Biophysical fragment screening of the beta1-adrenergic receptor: identification of high affinity arylpiperazine leads using structure-based drug design.* **J Med Chem**, 2013. 56(9): p. 3446-55.

180. Congreve, M.S., D.J. Davis, L. Devine, C. Granata, M. O'Reilly, P.G. Wyatt, and H. Jhoti, *Detection of ligands from a dynamic combinatorial library by X-ray crystallography*. **Angew Chem Int Ed Engl**, 2003. 42(37): p. 4479-82.

181. Ladbury, J.E., *Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design.* **Biochem Soc Trans**, 2010. 38(4): p. 888-93.

182. Kranz, J.K. and C. Schalk-Hihi, *Protein thermal shifts to identify low molecular weight fragments.* **Methods Enzymol**, 2011. 493: p. 277-98.

183. Lewis, L.M., L.J. Engle, W.E. Pierceall, D.E. Hughes, and K.J. Shaw, *Affinity capillary electrophoresis for the screening of novel antimicrobial targets.* **J Biomol Screen**, 2004. 9(4): p. 303-8.

184. Duong-Thi, M.D., E. Meiby, M. Bergstrom, T. Fex, R. Isaksson, and S. Ohlson, *Weak affinity chromatography as a new approach for fragment screening in drug discovery.* **Anal Biochem**, 2011. 414(1): p. 138-46.

185. Proll, F., P. Fechner, and G. Proll, *Direct optical detection in fragment-based screening.* **Anal Bioanal Chem**, 2009. 393(6-7): p. 1557-62.

186. Price, W.S., K. Hayamizu, H. Ide, and Y. Arata, *Strategies for diagnosing and alleviating artifactual attenuation associated with large gradient pulses in PGSE NMR diffusion measurements.* **J Magn Reson**, 1999. 139(2): p. 205-12.

187. Schanda, P., E. Kupce, and B. Brutscher, *SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds.* **J Biomol NMR**, 2005. 33(4): p. 199-211.

188. Schanda, P. and B. Brutscher, *Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds.* **J Am Chem Soc**, 2005. 127(22): p. 8014-5.

189. Pervushin, K., R. Riek, G. Wider, and K. Wuthrich, *Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.* **Proc Natl Acad Sci U S A**, 1997. 94(23): p. 12366-71.

190. Kay, L.E., M. Ikura, R. Tschudin, and A. Bax, *Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990.* **J Magn Reson**, 2011. 213(2): p. 423-41.

191. Bax, A. and M. Ikura, *An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins.* **J Biomol NMR**, 1991. 1(1): p. 99-104.

192. Grzesiek, S., H. Dobeli, R. Gentz, G. Garotta, A.M. Labhardt, and A. Bax, *1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma.* **Biochemistry**, 1992. 31(35): p. 8180-90.

193. Mayer, M. and B. Meyer, *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor.* **J Am Chem Soc**, 2001. 123(25): p. 6108-17.

194. Viegas, A., J. Manso, M.C. Corvo, M.M. Marques, and E.J. Cabrita, *Binding of ibuprofen, ketorolac, and diclofenac to COX-1 and COX-2 studied by saturation transfer difference NMR.* **J Med Chem**, 2011. 54(24): p. 8555-62.

195. Dalvit, C., G. Fogliatto, A. Stewart, M. Veronesi, and B. Stockman, *WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability.* **J Biomol NMR**, 2001. 21(4): p. 349-59.

196. Homola, J., *Present and future of surface plasmon resonance biosensors.* **Anal Bioanal Chem**, 2003. 377(3): p. 528-39.

197. Kelly, S.M. and N.C. Price, *The use of circular dichroism in the investigation of protein structure and function.* **Curr Protein Pept Sci**, 2000. 1(4): p. 349-84.

198. Whitmore, L. and B.A. Wallace, *DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data.* **Nucleic Acids Res**, 2004. 32(Web Server issue): p. W668-73.

199. Schuck, P. and P. Rossmanith, *Determination of the sedimentation coefficient distribution by least-squares boundary modeling.* **Biopolymers**, 2000. 54(5): p. 328-41.

200. Brown, P.H. and P. Schuck, *Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation.* **Biophys J**, 2006. 90(12): p. 4651-61.

201. Lobley, A., L. Whitmore, and B.A. Wallace, *DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra.* **Bioinformatics**, 2002. 18(1): p. 211-2.

202. Gradwell, M.J., H. Kogelberg, and T.A. Frenkiel, *Applying excitation sculpting to construct singly and doubly selective 1D NMR experiments.* **J Magn Reson**, 1997. 124(1): p. 267-70.

203. Loria, J.P., M. Rance, and A.G. Palmer, 3rd, *Transverse-relaxation-optimized (TROSY) gradient-enhanced triple-resonance NMR spectroscopy.* **J Magn Reson**, 1999. 141(1): p. 180-4.

204. Adams, P.D., R.W. Grosse-Kunstleve, L.W. Hung, T.R. Ioerger, A.J. McCoy, N.W. Moriarty, R.J. Read, J.C. Sacchettini, N.K. Sauter, and T.C. Terwilliger, *PHENIX: building new software for automated crystallographic structure determination.* **Acta Crystallogr D Biol Crystallogr**, 2002. 58(Pt 11): p. 1948-54.

205. Kim, K.A., J.S. Song, J. Jee, M.R. Sheen, C. Lee, T.G. Lee, S. Ro, J.M. Cho, W. Lee, T. Yamazaki, Y.H. Jeon, and C. Cheong, *Structure of human PRL-3, the phosphatase associated with cancer metastasis.* **FEBS Lett**, 2004. 565(1-3): p. 181-7.

206. Meier, S., Y.C. Li, J. Koehn, I. Vlattas, J. Wareing, W. Jahnke, L.P. Wennogle, and S. Grzesiek, *Backbone resonance assignment of the 298 amino acid catalytic domain of protein tyrosine phosphatase 1B (PTP1B).* **J Biomol NMR**, 2002. 24(2): p. 165-6.

207. Francis, D.M., R. Page, and W. Peti, *Sequence-specific backbone (1)H, (1)(3)C and (1)(5)N assignments of the 34 kDa catalytic domain of PTPN5 (STEP).* **Biomol NMR Assign**, 2014. 8(1): p. 185-8.

208. Gordon, J.A., *Use of vanadate as protein-phosphotyrosine phosphatase inhibitor.* **Methods Enzymol**, 1991. 201: p. 477-82.

209. Sattler, M. and S.W. Fesik, *Use of deuterium labeling in NMR: overcoming a sizeable problem.* **Structure**, 1996. 4(11): p. 1245-9.

210. Huang, X., I.M. de Vera, A.M. Veloro, M.E. Blackburn, J.L. Kear, J.D. Carter, J.R. Rocca, C. Simmerling, B.M. Dunn, and G.E. Fanucci, *Inhibitor-induced conformational shifts and ligand-exchange dynamics for HIV-1 protease measured by pulsed EPR and NMR spectroscopy.* **J Phys Chem B**, 2012. 116(49): p. 14235-44.

211. Jonsson, U., L. Fagerstam, B. Ivarsson, B. Johnsson, R. Karlsson, K. Lundh, S. Lofas, B. Persson, H. Roos, I. Ronnberg, and et al., *Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology.* **Biotechniques**, 1991. 11(5): p. 620-7.

212. Zhuo, S., J.C. Clemens, D.J. Hakes, D. Barford, and J.E. Dixon, *Expression, purification, crystallization, and biochemical characterization of a recombinant protein phosphatase.* **J Biol Chem**, 1993. 268(24): p. 17754-61.

213. Zhang, Z.Y., J.C. Clemens, H.L. Schubert, J.A. Stuckey, M.W. Fischer, D.M. Hume, M.A. Saper, and J.E. Dixon, *Expression, purification, and physicochemical characterization of a recombinant Yersinia protein tyrosine phosphatase.* **J Biol Chem**, 1992. 267(33): p. 23759-66.

214. Pot, D.A., T.A. Woodford, E. Remboutsika, R.S. Haun, and J.E. Dixon, *Cloning, bacterial expression, purification, and characterization of the cytoplasmic domain of rat LAR, a receptor-like protein tyrosine phosphatase.* **J Biol Chem**, 1991. 266(29): p. 19688-96.

215. Takai, A. and G. Mieskes, *Inhibitory effect of okadaic acid on the p-nitrophenyl phosphate phosphatase activity of protein phosphatases.* **Biochem J**, 1991. 275 ( Pt 1): p. 233-9.

216. McGovern, S.L., B.T. Helfand, B. Feng, and B.K. Shoichet, *A specific mechanism of nonspecific inhibition.* **J Med Chem**, 2003. 46(20): p. 4265-72.

217. Feng, B.Y. and B.K. Shoichet, *A detergent-based assay for the detection of promiscuous inhibitors.* **Nat Protoc**, 2006. 1(2): p. 550-3.

218. Ryan, A.J., N.M. Gray, P.N. Lowe, and C.W. Chung, *Effect of detergent on "promiscuous" inhibitors.* **J Med Chem**, 2003. 46(16): p. 3448-51.

219. Pot, D.A. and J.E. Dixon, *Active site labeling of a receptor-like protein tyrosine phosphatase.* **J Biol Chem**, 1992. 267(1): p. 140-3.

220. Cronan, J.E., Jr., *Biotination of proteins in vivo. A post-translational modification to label, purify, and study proteins.* **J Biol Chem**, 1990. 265(18): p. 10327-33.

221. Navratilova, I. and A.L. Hopkins, *Fragment screening by surface plasmon resonance.* **ACS Med Chem Lett**, 2010. 1(1): p. 44-8.

222. Hajduk, P.J., J.R. Huth, and S.W. Fesik, *Druggability indices for protein targets derived from NMR-based screening data.* **J Med Chem**, 2005. 48(7): p. 2518-25.

223. Hajduk, P.J., J.R. Huth, and C. Tse, *Predicting protein druggability.* **Drug Discov Today**, 2005. 10(23-24): p. 1675-82.

224. Davies, T.G. and I.J. Tickle, *Fragment screening using X-ray crystallography.* **Top Curr Chem**, 2012. 317: p. 33-59.

225. Jhoti, H., A. Cleasby, M. Verdonk, and G. Williams, *Fragment-based screening using X-ray crystallography and NMR spectroscopy.* **Curr Opin Chem Biol**, 2007. 11(5): p. 485-93.

226. Stockman, B.J., M. Kothe, D. Kohls, L. Weibley, B.J. Connolly, A.L. Sheils, Q. Cao, A.C. Cheng, L. Yang, A.V. Kamath, Y.H. Ding, and M.E. Charlton, *Identification of allosteric PIF-pocket ligands for PDK1 using NMR-based fragment screening and 1H-15N TROSY experiments.* **Chem Biol Drug Des**, 2009. 73(2): p. 179-88.

227. Wielens, J., S.J. Headey, D.I. Rhodes, R.J. Mulder, O. Dolezal, J.J. Deadman, J. Newman, D.K. Chalmers, M.W. Parker, T.S. Peat, and M.J. Scanlon, *Parallel screening of low molecular weight fragment libraries: do differences in methodology affect hit identification?* **J Biomol Screen**, 2013. 18(2): p. 147-59.

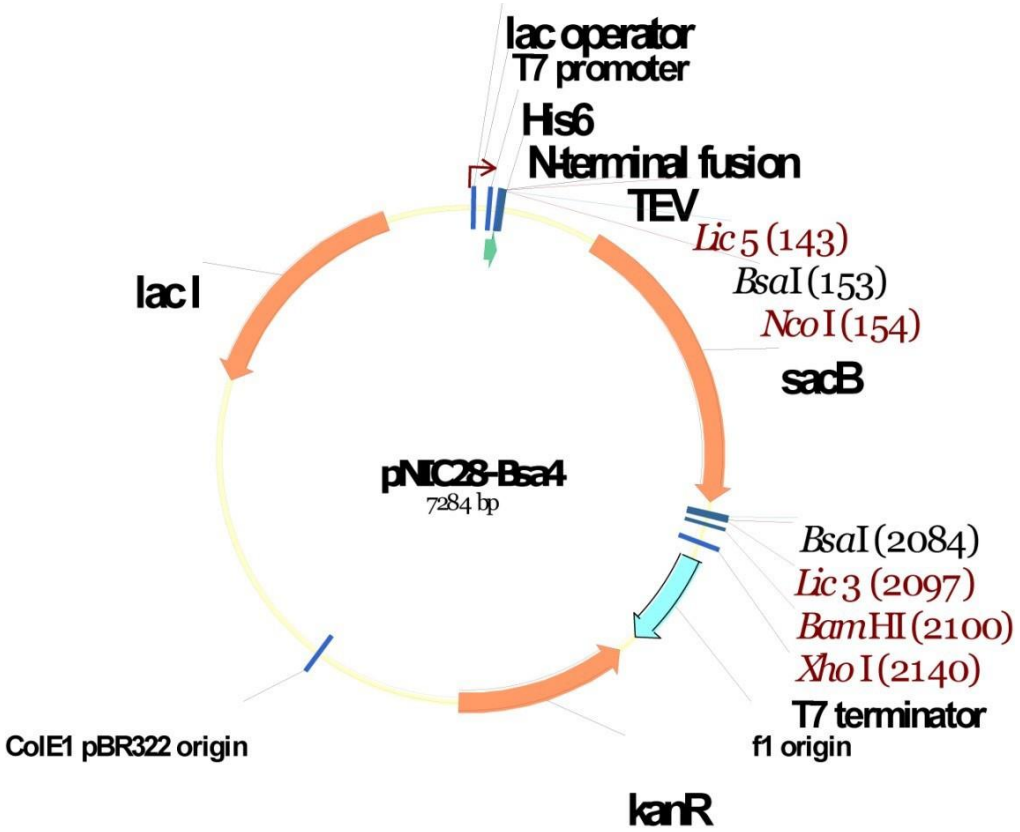
228. Ericsson, U.B., B.M. Hallberg, G.T. Detitta, N. Dekker, and P. Nordlund, *Thermofluor-based high-throughput stability optimization of proteins for structural studies.* **Anal Biochem**, 2006. 357(2): p. 289-98.

229. Ward, R.A., C. Brassington, A.L. Breeze, A. Caputo, S. Critchlow, G. Davies, L. Goodwin, G. Hassall, R. Greenwood, G.A. Holdgate, M. Mrosek, R.A. Norman, S. Pearson, J. Tart, J.A. Tucker, M. Vogtherr, D. Whittaker, J. Wingfield, J. Winter, and K. Hudson, *Design and synthesis of novel lactate dehydrogenase A inhibitors by fragment-based lead generation.* **J Med Chem**, 2012. 55(7): p. 3285-306.

230. He, R., Z.H. Yu, R.Y. Zhang, L. Wu, A.M. Gunawan, B.S. Lane, J.S. Shim, L.F. Zeng, Y. He, L. Chen, C.D. Wells, J.O. Liu, and Z.Y. Zhang, *Exploring the Existing Drug Space for Novel pTyr Mimetic and SHP2 Inhibitors.* **ACS Med Chern Lett,** 2015. 6(7): p. 782-6.

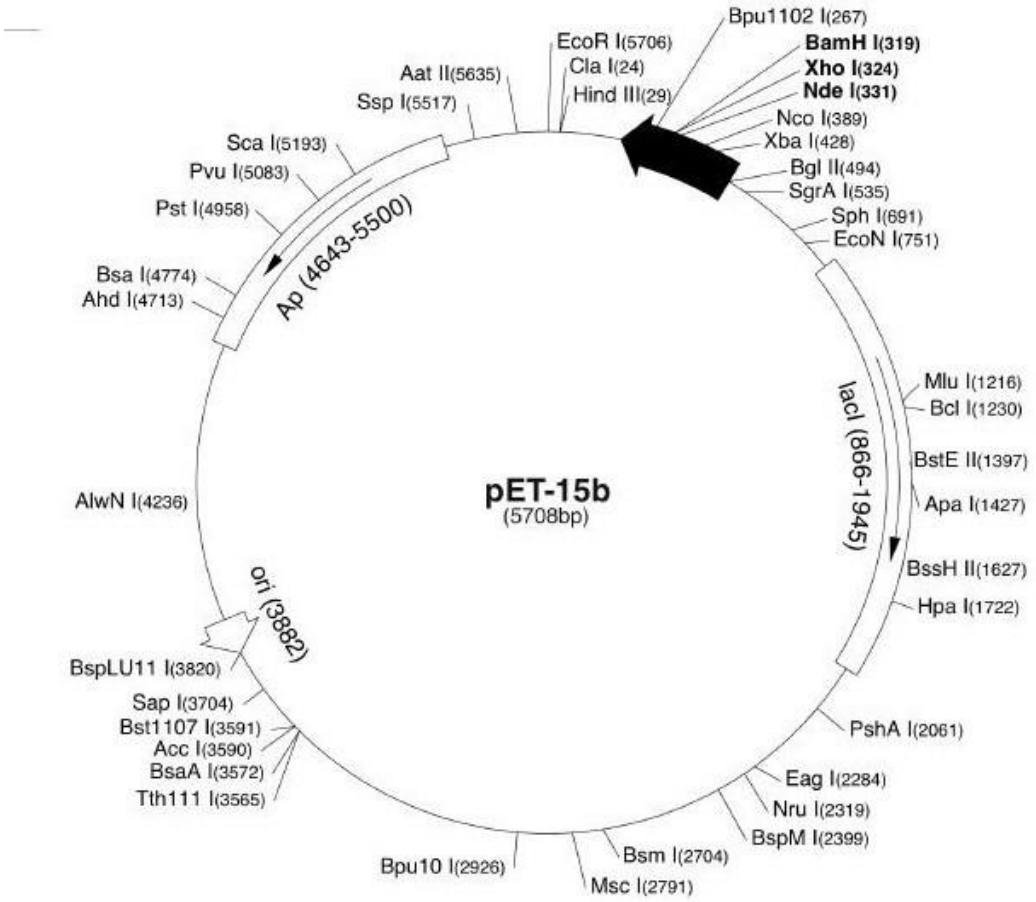
# Appendix

**Figure 1a – vector map of pNIC-Bsa4**



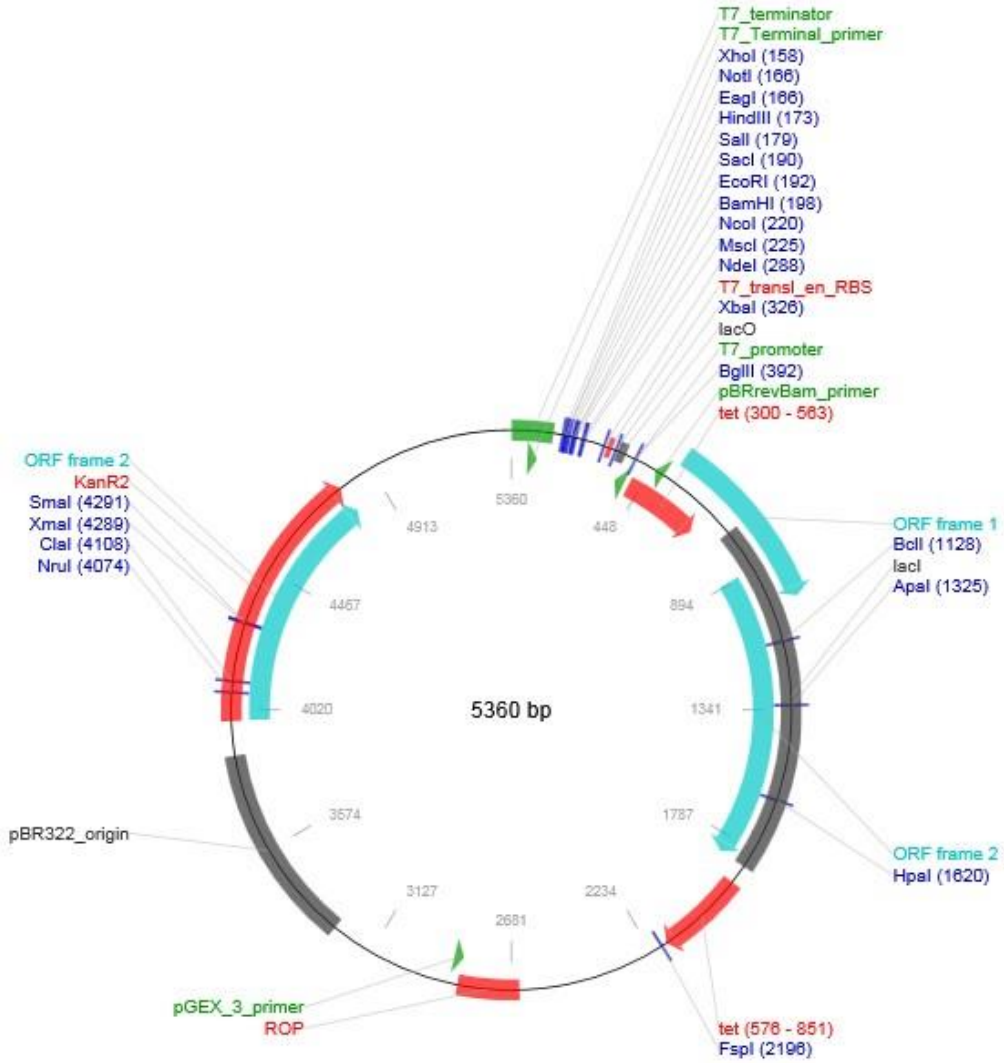

The pNIC28-Bsa4 vector was a kind gift from Dr Alastair Barr and Professor Stefan Knapp at the Structural Genomics Consortium, University of Oxford. The pNIC28-Bsa4 vector is a pET expression system containing a $His_{x6}$ affinity tag as part of a 22 amimo-acid linker which harbours a TEV protease cleavage site. The pNIC28-Bsa4 vector possesses a Kanamycin resistance gene and gene expression is under the control of a T7 promoter system.

**Appendix figure 1b –vector map of pET15b**




The pET-15b vector carries an N-terminal $His_{x6}$ tag sequence immediately followed by a thrombin site and three cloning sites, and is under the control of a T7 promoter.

**Appendix figure 1c –vector map of pET-26b(+)**




pET-26b(+) carries an N-terminal *pelB* signal sequence (which acts as a periplasmic localisation signal) and also contains a non-cleavable C-terminal $His_{x6}$ tag

**Appendix figure 2 –SHP2 DNA constructs**

**Summary of SHP2 constructs used in this study:**

| Protein construct | Vector |
|---|---|
| **1. SHP2 catalytic domain (aa residues 237-533)** | pNIC28-Bsa4 |
| **2. SHP2 Tandem SH2 domain (aa residues 4-216)** | pNIC28-Bsa4 |
| **3. Full length SHP2 (aa residues 1-597)** | pET-26b(+) |
| **4. SHP2 catalytic domain (SPR) (aa residues 237-533)** | pET-15b(+) |
| **5. SHP2 Tandem SH2 domain (SPR) (aa residues 4-216)** | pET-15b(+) |
| **6. Full length SHP2 (SPR) (aa residues 1-597)** | pET-15b(+) |

**DNA sequences for each of these constructs including the SHP2 gene insertion and corresponding cloning sites are provided below.**

# Shp2 catalytic domain

## Vector: pNIC-Bsa4

Shp2 catalytic domain, N-terminal $His_{x6}$ tag (plus cleavage site and linker region)

## Protein sequence

MGHHHHHHSSGVDLGTENLYFQ*SMAETTDKVKQGFWEEFETLQQQECKLLYSRKEGQRQENKNKN
RYKNILPFDHTRVVLHDGDPNEPVSDYINANIIMPEFETKCNNSKPKKSYIATQGCLQNTVNDFWR
MVFQENSRVIVMTTKEVERGKSKCVKYWPDEYALKEYGVMRVRNVKESAAHDYTLRELKLSKVGQG
NTERTVWQYHFRTWPDHGVPSDPGGVLDFLEEVHHKQESIMDAGPVVVHCSAGIGRTGTFIVIDIL
IDIIREKGVDCDIDVPKTIQMVRSQRSGMVQTEAQYRFIYMAVQHYIETLQRRI

MGHHHHHHSSGVDLGTENLYFQ* - denotes TEV-cleavable $His_{x6}$ tag

## DNA sequence

Cloning site: NdeI – XhoI

5’CCATGG

atgggccatcatcatcatcatagcagcggcgtggatctgggcaccgaaaacctgtat
tttcagagcatggcggaaaccaccgataaagtgaaacagggcttttgggaagaatttgaa
accctgcagcagcaggaatgcaaactgctgtatagccgcaaagaaggccagcgccaggaa
aacaaaaacaaaaaccgctataaaaacattctgccgtttgatcatacccgcgtggtgctg
catgatggcgatccgaacgaaccggtgagcgattatattaacgcgaacattattatgccg
gaatttgaaaccaaatgcaacaacagcaaaccgaaaaaaagctatattgcgacccagggc
tgcctgcagaacaccgtgaacgatttttggcgcatggtgtttcaggaaaacagccgcgtg
attgtgatgaccaccaaagaagtggaacgcggcaaaagcaaatgcgtgaaatattggccg
gatgaatatgcgctgaaagaatatggcgtgatgcgcgtgcgcaacgtgaaagaaagcgcg
gcgcatgattataccctgcgcgaactgaaactgagcaaagtgggccagggcaacaccgaa
cgcaccgtgtggcagtatcattttcgcacctggccggatcatggcgtgccgagcgatccg
ggcggcgtgctggattttctggaagaagtgcatcataaacaggaaagcattatggatgcg
ggcccggtggtggtgcattgcagcgcgggcattggccgcaccggcacctttattgtgatt
gatattctgattgatattattcgcgaaaaaggcgtggattgcgatattgatgtgccgaaa
accattcagatggtgcgcagccagcgcagcggcatggtgcagaccgaagcgcagtatcgc
tttatttatatggcggtgcagcattatattgaaaccctgcagcgccgcatt

3’ CTCGAG

# Tandem SH2 domain

## Vector: pNIC-Bsa4

Tandem SH2 domain, N-terminal $His_{x6}$ tag (plus cleavage site and linker region)

## Protein sequence

MHHHHHHSSGVDLGTENLYFQ*SMRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLSV
RRNGAVTHIKIQNTGDYYDLYGGEKFATLAELVQYYMEHHGQLKEKNGDVIELKYPLNCADPTSE
RWFHGHLSGKEAEKLLTEKGKHGSFLVRESQSHPGDFVLSVRTGDDKGESNDGKSKVTHVMIRCQ
ELKYDVGGGERFDSLTDLVEHYKKNPMVETLGTVLQLKQPL

MGHHHHHHSSGVDLGTENLYFQ* - denotes TEV-cleavable $His_{x6}$ tag

## DNA sequence

Cloning site: NdeI – XhoI

5’CCATGG

atgcatcatcatcatcatagcagcggcgtggatctgggcaccgaaaacctgtatttt
cagagcatgcgccgctggtttcatccgaacattaccggcgtggaagcggaaaacctgctg
ctgacccgcggcgtggatggcagctttctggcgcgcccgagcaaaagcaacccgggcgat
tttaccctgagcgtgcgccgcaacggcgcggtgacccatattaaaattcagaacaccggc
gattattatgatctgtatggcggcgaaaaatttgcgaccctggcggaactggtgcagtat
tatatggaacatcatggccagctgaaagaaaaaaacggcgatgtgattgaactgaaatat
ccgctgaactgcgcggatccgaccagcgaacgctggtttcatggccatctgagcggcaaa
gaagcggaaaaactgctgaccgaaaaaggcaaacatggcagctttctggtgcgcgaaagc
cagagccatccgggcgattttgtgctgagcgtgcgcaccggcgatgataaaggcgaaagc
aacgatggcaaaagcaaagtgacccatgtgatgattcgctgccaggaactgaaatatgat
gtgggcggcggcgaacgctttgatagcctgaccgatctggtggaacattataaaaaaaac
ccgatggtggaaaccctgggcaccgtgctgcagctgaaacagccgctg

3’ CTCGAG

# Full length SHP2

**Vector: pET-26B(+)**

Full length SHP2, non-cleavable C-terminal $His_{x6}$ tag (pre-cloned into pET-26B(+)

**Protein sequence**

MTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLSVRRNGAVTHIKIQNTGDYYDLY
GGEKFATLAELVQYYMEHHGQLKEKNGDVIELKYPLNCADPTSERWFHGHLSGKEAEKLLTEKGKH
GSFLVRESQSHPGDFVLSVRTGDDKGESNDGKSKVTHVMIRCQELKYDVGGGERFDSLTDLVEHYK
KNPMVETLGTVLQLKQPLNTTRINAAEIESRVRELSKLAETTDKVKQGFWEEFETLQQQECKLLYS
RKEGQRQENKNKNRYKNILPFDHTRVVLHDGDPNEPVSDYINANIIMPEFETKCNNSKPKKSYIAT
QGCLQNTVNDFWRMVFQENSRVIVMTTKEVERGKSKCVKYWPDEYALKEYGVMRVRNVKESAAHDY
TLRELKLSKVGQGNTERTVWQYHFRTWPDHGVPSDPGGVLDFLEEVHHKQESIMDAGPVVVHCSAG
IGRTGTFIVIDILIDIIREKGVDCDIDVPKTIQMVRSQRSGMVQTEAQYRFIYMAVQHYIETLQRR
IEEEQKSKRKGHEYTNIKYSLADQTSGDQSPLPPCTPTPPCAEMREDSARVYENVGLMQQQKSFR

**DNA sequence**

Cloning site: NdeI – XhoI

5’CCATGG

atgaccagccgccgctggtttcatccgaacattaccggcgtggaagcggaaaacctgctgctgacc
cgcggcgtggatggcagctttctggcgcgcccgagcaaaagcaacccgggcgattttaccctgagc
gtgcgccgcaacggcgcggtgacccatattaaaattcagaacaccggcgattattatgatctgtat
ggcggcgaaaaatttgcgaccctggcggaactggtgcagtattatatggaacatcatggccagctg
aaagaaaaaaacggcgatgtgattgaactgaaatatccgctgaactgcgcggatccgaccagcgaa
cgctggtttcatggccatctgagcggcaaagaagcggaaaaactgctgaccgaaaaaggcaaacat
ggcagctttctggtgcgcgaaagccagagccatccgggcgattttgtgctgagcgtgcgcaccggc
gatgataaaggcgaaagcaacgatggcaaaagcaaagtgacccatgtgatgattcgctgccaggaa
ctgaaatatgatgtgggcggcggcgaacgctttgatagcctgaccgatctggtggaacattataaa
aaaaacccgatggtggaaaccctgggcaccgtgctgcagctgaaacagccgctgaacaccacccgc
attaacgcggcggaaattgaaagccgcgtgcgcgaactgagcaaactggcggaaaccaccgataaa
gtgaaacagggcttttgggaagaatttgaaaccctgcagcagcaggaatgcaaactgctgtatagc
cgcaaagaaggccagcgccaggaaaacaaaaacaaaaaccgctataaaaacattctgccgtttgat
catacccgcgtggtgctgcatgatggcgatccgaacgaaccggtgagcgattatattaacgcgaac
attattatgccggaatttgaaaccaaatgcaacaacagcaaaccgaaaaaaagctatattgcgacc
cagggctgcctgcagaacaccgtgaacgatttttggcgcatggtgtttcaggaaaacagccgcgtg
attgtgatgaccaccaaagaagtggaacgcggcaaaagcaaatgcgtgaaatattggccggatgaa
tatgcgctgaaagaatatggcgtgatgcgcgtgcgcaacgtgaaagaaagcgcggcgcatgattat
accctgcgcgaactgaaactgagcaaagtgggccagggcaacaccgaacgcaccgtgtggcagtat
cattttcgcacctggccggatcatggcgtgccgagcgatccgggcggcgtgctggattttctggaa
gaagtgcatcataaacaggaaagcattatggatgcgggcccggtggtggtgcattgcagcgcgggc
attggccgcaccggcacctttattgtgattgatattctgattgatattattcgcgaaaaaggcgtg

gattgcgatattgatgtgccgaaaaccattcagatggtgcgcagccagcgcagcggcatggtgcag
accgaagcgcagtatcgctttatttatatggcggtgcagcattatattgaaaccctgcagcgccgc
attgaagaagaacagaaaagcaaacgcaaaggccatgaatataccaacattaaatatagcctggcg
gatcagaccagcggcgatcagagcccgctgccgccgtgcaccccgaccccgccgtgcgcggaaatg
cgcgaagatagcgcgcgcgtgtatgaaaacgtgggcctgatgcagcagcagaaaagctttcgc

3’ CTCGAG

# Shp2 catalytic domain (SPR)

## Vector: pET-15b(+)

Shp2 catalytic domain, N-terminal $His_{x6}$ tag (plus cleavage site and linker region) and C-terminal (Avi-Tag) biotinylation sequence

## Protein sequence

MGHHHHHHSSGVDLGTENLYFQ*SMAETTDKVKQGFWEEFETLQQQECKLLYSRKEGQRQENKNKN
RYKNILPFDHTRVVLHDGDPNEPVSDYINANIIMPEFETKCNNSKPKKSYIATQGCLQNTVNDFWR
MVFQENSRVIVMTTKEVERGKSKCVKYWPDEYALKEYGVMRVRNVKESAAHDYTLRELKLSKVGQG
NTERTVWQYHFRTWPDHGVPSDPGGVLDFLEEVHHKQESIMDAGPVVVHCSAGIGRTGTFIVIDIL
IDIIREKGVDCDIDVPKTIQMVRSQRSGMVQTEAQYRFIYMAVQHYIETLQRRIGLNDIFEAQKIE
WHE*

MGHHHHHHSSGVDLGTENLYFQ* - denotes TEV-cleavable $His_{x6}$ tag

GLNDIFEAQKIEWHE* - denotes C-terminal Avi-tag biotinylation sequence

## DNA sequence

Cloning site: NdeI – XhoI

5’CCATGG

ccatgggccatcatcatcatcatcatagcagcggcgtggatctgggcaccgaaaacctgtattttc
agagcatggcggaaaccaccgataaagtgaaacagggcttttgggaagaatttgaaaccctgcagc
agcaggaatgcaaactgctgtatagccgcaaagaaggccagcgccaggaaaacaaaaacaaaaacc
gctataaaaacattctgccgtttgatcatacccgcgtggtgctgcatgatggcgatccgaacgaac
cggtgagcgattatattaacgcgaacattattatgccggaatttgaaaccaaatgcaacaacagca
aaccgaaaaaaagctatattgcgacccagggctgcctgcagaacaccgtgaacgatttttggcgca
tggtgtttcaggaaaacagccgcgtgattgtgatgaccaccaaagaagtggaacgcggcaaaagca
aatgcgtgaaatattggccggatgaatatgcgctgaaagaatatggcgtgatgcgcgtgcgcaacg
tgaaagaaagcgcggcgcatgattataccctgcgcgaactgaaactgagcaaagtgggccagggca
acaccgaacgcaccgtgtggcagtatcattttcgcacctggccggatcatggcgtgccgagcgatc
cgggcggcgtgctggattttctggaagaagtgcatcataaacaggaaagcattatggatgcgggcc
cggtggtggtgcattgcagcgcgggcattggccgcaccggcacctttattgtgattgatattctga
ttgatattattcgcgaaaaaggcgtggattgcgatattgatgtgccgaaaaccattcagatggtgc
gcagccagcgcagcggcatggtgcagaccgaagcgcagtatcgctttatttatatggcggtgcagc

attatattgaaaccctgcagcgccgcattggcctgaacgatattttgaagcgcagaaaattgaat
ggcatgaataa

3’ CTCGAG

# Tan-SH2 domain (SPR)

**Vector: pET-15b(+)**

Tan-SH2 domain, N terminal $His_{x6}$ tag (plus cleavage site and linker region) and C-terminal (Avi-Tag) biotinylation sequence

**Protein sequence**

MHHHHHHSSGVDLGTENLYFQ*MTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLS
VRRNGAVTHIKIQNTGDYYDLYGGEKFATLAELVQYYMEHHGQLKEKNGDVIELKYPLNCADPTSE
RWFHGHLSGKEAEKLLTEKGKHGSFLVRESQSHPGDFVLSVRTGDDKGESNDGKSKVTHVMIRCQE
LKYDVGGGERFDSLTDLVEHYKKNPMVETLGTVLQLKQPLGLNDIFEAQKIEWHE*

MGHHHHHHSSGVDLGTENLYFQ* - denotes TEV-cleavable $His_{x6}$ tag

GLNDIFEAQKIEWHE* - denotes C-terminal Avi-tag biotinylation sequence

**DNA sequence**

Cloning site: NdeI – XhoI

5’CCATGG

GGCCATCACCATCACCATCACTCATCGGGCGTTGACCTGGGCACCGAAAATCTGTACTTCCAAATG
ACCTCGCGTCGTTGGTTCCACCCGAACATTACCGGTGTCGAAGCAGAAAATCTGCTGCTGACGCGT
GGTGTGGATGGCAGTTTTCTGGCTCGCCCGAGCAAATCTAACCCGGGTGACTTCACCCTGTCCGTT
CGTCGCAACGGCGCGGTCACCCATATTAAAATCCAGAATACGGGCGATTATTACGACCTGTATGGC
GGTGAAAAATTTGCGACCCTGGCCGAACTGGTGCAGTATTACATGGAACATCACGGTCAACTGAAA
GAGAAAAACGGCGATGTTATTGAACTGAAATACCCGCTGAATTGCGCGGACCCGACCAGCGAACGT
TGGTTTCATGGCCACCTGTCTGGCAAAGAAGCCGAAAAACTGCTGACGGAAAAAGGTAAACATGGC
AGCTTTCTGGTGCGTGAAAGTCAGTCCCACCCGGGTGATTTCGTGCTGTCTGTTCGCACCGGTGAT
GACAAAGGCGAATCAAACGATGGCAAATCGAAAGTCACGCATGTGATGATCCGTTGTCAAGAACTG
AAATATGATGTTGGCGGTGGCGAACGCTTTGATAGCCTGACCGACCTGGTCGAACACTACAAGAAA
AACCCGATGGTGGAAACCCTGGGTACGGTTCTGCAACTGAAACAACCGCTGGGCCTGAATGACATC
TTTGAAGCACAGAAAATTGAATGGCACGAATGA

3’ CTCGAG

# Full length SHP2(SPR)

**Vector: pET-15b(+)**

Full length SHP2, N-terminal $His_{x6}$ tag (plus cleavage site and linker region) and C terminal (Avi-Tag) biotinylation sequence

**Protein sequence**

MHHHHHHSSGVDLGTENLYFQ*MTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLS
VRRNGAVTHIKIQNTGDYYDLYGGEKFATLAELVQYYMEHHGQLKEKNGDVIELKYPLNCADPTSE
RWFHGHLSGKEAEKLLTEKGKHGSFLVRESQSHPGDFVLSVRTGDDKGESNDGKSKVTHVMIRCQE
LKYDVGGGERFDSLTDLVEHYKKNPMVETLGTVLQLKQPLNTTRINAAEIESRVRELSKLAETTDK
VKQGFWEEFETLQQQECKLLYSRKEGQRQENKNKNRYKNILPFDHTRVVLHDGDPNEPVSDYINAN
IIMPEFETKCNNSKPKKSYIATQGCLQNTVNDFWRMVFQENSRVIVMTTKEVERGKSKCVKYWPDE
YALKEYGVMRVRNVKESAAHDYTLRELKLSKVGQALLQGNTERTVWQYHFRTWPDHGVPSDPGGVL
DFLEEVHHKQESIMDAGPVVVHCSAGIGRTGTFIVIDILIDIIREKGVDCDIDVPKTIQMVRSQRS
GMVQTEAQYRFIYMAVQHYIETLQRRIEEEQKSKRKGHEYTNIKYSLADQTSGDQSPLPPCTPTPP
CAEMREDSARVYENVGLMQQQKSFRGLNDIFEAQKIEWHE

MGHHHHHHSSGVDLGTENLYFQ* - denotes TEV-cleavable $His_{x6}$ tag

GLNDIFEAQKIEWHE* - denotes C-terminal Avi-tag biotinylation sequence

**DNA sequence**

Cloning site: NdeI – XhoI

5’CCATGG

atgcatcatcatcatcatcatagcagcggcgtggatctgggcaccgaaaacctgtattttcagatg
accagccgccgctggtttcatccgaacattaccggcgtggaagcggaaaacctgctgctgacccgc
ggcgtggatggcagctttctggcgcgcccgagcaaaagcaacccgggcgattttaccctgagcgtg
cgccgcaacggcgcggtgacccatattaaaattcagaacaccggcgattattatgatctgtatggc
ggcgaaaaatttgcgaccctggcggaactggtgcagtattatatggaacatcatggccagctgaaa
gaaaaaaacggcgatgtgattgaactgaaatatccgctgaactgcgcggatccgaccagcgaacgc
tggtttcatggccatctgagcggcaaagaagcggaaaaactgctgaccgaaaaaggcaaacatggc
agctttctggtgcgcgaaagccagagccatccgggcgattttgtgctgagcgtgcgcaccggcgat
gataaaggcgaaagcaacgatggcaaaagcaaagtgacccatgtgatgattcgctgccaggaactg
aaatatgatgtgggcggcggcgaacgctttgatagcctgaccgatctggtggaacattataaaaaa
aacccgatggtggaaaccctgggcaccgtgctgcagctgaaacagccgctgaacaccacccgcatt
aacgcggcggaaattgaaagccgcgtgcgcgaactgagcaaactggcggaaaccaccgataaagtg
aaacagggcttttgggaagaatttgaaaccctgcagcagcaggaatgcaaactgctgtatagccgc
aaagaaggccagcgccaggaaaacaaaaacaaaaaccgctataaaaacattctgccgtttgatcat
acccgcgtggtgctgcatgatggcgatccgaacgaaccggtgagcgattatattaacgcgaacatt
attatgccggaatttgaaaccaaatgcaacaacagcaaaccgaaaaaaagctatattgcgacccag
ggctgcctgcagaacaccgtgaacgatttttggcgcatggtgtttcaggaaaacagccgcgtgatt
gtgatgaccaccaaagaagtggaacgcggcaaaagcaaatgcgtgaaatattggccggatgaatat
gcgctgaaagaatatggcgtgatgcgcgtgcgcaacgtgaaagaaagcgcggcgcatgattatacc
ctgcgcgaactgaaactgagcaaagtgggccaggcgctgctgcagggcaacaccgaacgcaccgtg
tggcagtatcattttcgcacctggccggatcatggcgtgccgagcgatccgggcggcgtgctggat
tttctggaagaagtgcatcataaacaggaaagcattatggatgcgggcccggtggtggtgcattgc
agcgcgggcattggccgcaccggcacctttattgtgattgatattctgattgatattattcgcgaa

aaaggcgtggattgcgatattgatgtgccgaaaaccattcagatggtgcgcagccagcgcagcggc
atggtgcagaccgaagcgcagtatcgctttatttatatggcggtgcagcattatattgaaaccctg
cagcgccgcattgaagaagaacagaaaagcaaacgcaaaggccatgaatataccaacattaaatat
agcctggcggatcagaccagcggcgatcagagcccgctgccgccgtgcaccccgaccccgccgtgc
gcggaaatgcgcgaagatagcgcgcgcgtgtatgaaaacgtgggcctgatgcagcagcagaaaagc
tttcgcggcctgaacgatatttttgaagcgcagaaaattgaatggcatga

3' CTCGAG

**Appendix figure 3a: Maybridge Fragment Library Collection**

| CHEMICAL NAME | SMIILES | BARCODE |
|---|---|---|
| 2-ethyl-4-methyl-1H-imidazole | CCc1[nH]cc(n1)C | 0103512998 |
| 4-methylquinolin-2-ol | CC1=CC(=NC2=C1C=CC=C2)O | 0103512997 |
| [(5-methoxy-1H-indol-3-yl)methyl]dimethylamine | COC1=CC2=C([NH]C=C2CN(C)C)C=C1 | 0103512996 |
| 7-methyl-1H-indole | CC1=C2[NH]C=CC2=CC=C1 | 0103512995 |
| 2-(pyridin-3-yl)acetic acid hydrochloride | Cl.OC(=O)CC1=CC=CN=C1 | 0103512994 |
| 1-benzylpiperidin-3-ol hydrochloride | Cl.OC1CCCN(C1)CC2=CC=CC=C2 | 0103512993 |
| 4-(1H-imidazol-1-yl)phenol | OC1=CC=C(C=C1)[N]2C=CN=C2 | 0103512992 |
| 3-phenyl-1H-pyrazole | c1ccc(cc1)c2cc[nH]n2 | 0103512991 |
| 6-chloro-1,3-benzoxazol-2-ol | OC1=NC2=C(O1)C=C(Cl)C=C2 | 0103512990 |
| 3-methyl-2,3-dihydro-1,3-benzoxazole-2-thione | CN1C(=S)OC2=C1C=CC=C2 | 0103512989 |
| 1H-indazol-5-amine | Nc1ccc2c(c1)cn[nH]2 | 0103512119 |
| methyl 1H-indole-3-carboxylate | COC(=O)c1c[nH]c2c1cccc2 | 0103512120 |
| 2-methylfuran-3-carboxylic acid | Cc1c(cco1)C(=O)O | 0103512975 |
| 2-methyl-1H-indol-5-amine | Cc1cc2cc(ccc2[nH]1)N | 0103512976 |
| 1,3-benzothiazol-2-ol | c1ccc2c(c1)nc(s2)O | 0103512977 |
| ethyl 2-amino-4-(4-fluorophenyl)thiophene-3-carboxylate | CCOC(=O)c1c(N)scc1-c1ccc(F)cc1 | 0103512978 |
| 1-[4-(trifluoromethyl)pyrimidin-2-yl]piperidin-4-ol | OC1CCN(CC1)C2=NC(=CC=N2)C(F)(F)F | 0103512979 |
| 4-(2-methyl-1,3-thiazol-4-yl)aniline | Cc1nc(cs1)-c1ccc(N)cc1 | 0103512980 |
| 2-amino-5-phenylfuran-3-carbonitrile | NC1=C(C=C(O1)C2=CC=CC=C2)C#N | 0103512981 |
| 8-methyl-1,2,3,4-tetrahydroquinoline | Cc1cccc2CCCNc12 | 0103512982 |
| 4-(morpholin-4-yl)aniline | c1cc(ccc1N)N2CCOCC2 | 0103512983 |

| | | |
|---|---|---|
| 2,3,4,5-tetrahydro-1H-1,4-benzodiazepine-2,5-dione | O=C1CNC(=O)C2=CC=CC=C2N1 | 0103512984 |
| 2-(morpholin-4-yl)aniline | NC1=C(C=CC=C1)N2CCOCC2 | 0103512985 |
| 5-amino-1-phenyl-1H-pyrazole-4-carboxamide | c1ccc(cc1)n2c(c(cn2)C(=O)N)N | 0103512986 |
| 3-(2-methoxyphenyl)propanoic acid | COc1ccccc1CCC(=O)[O-] | 0103512974 |
| 1-phenyl-1,2,3,6-tetrahydropyridazine-3,6-dione | O=c1ccc(=O)n([nH]1)c1ccccc1 | 0103512973 |
| 5-amino-2-(1H-pyrrol-1-yl)benzonitrile | c1ccn(c1)c2ccc(cc2C#N)N | 0103512972 |
| 5-methyl-2-phenyl-2H-1,2,3-triazole-4-carboxamide | Cc1c(nn(n1)c2ccccc2)C(=O)N | 0103512971 |
| 3-chloro-4-fluorobenzamide | c1cc(c(cc1C(=O)N)Cl)F | 0103512970 |
| 6-(2,6-dichlorophenoxy)pyridin-3-amine | c1cc(c(c(c1)Cl)Oc2ccc(cn2)N)Cl | 0103512969 |
| 2-(piperidin-1-yl)benzamide | c1ccc(c(c1)C(=O)N)N2CCCCC2 | 0103512968 |
| 5-fluoroquinazolin-4-ol | OC1=NC=NC2=C1C(=CC=C2)F | 0103512967 |
| N-(4-propylphenyl)acetamide | CCCC1=CC=C(C=C1)NC(=O)C | 0103512966 |
| 6-methyl-1,3-benzothiazol-2-amine | Cc1ccc2c(c1)sc(n2)N | 0103512965 |
| 6-chloro-1,3-benzothiazol-2-amine | c1cc2c(cc1Cl)sc(n2)N | 0103512964 |
| 3,5-dichlorobenzene-1-sulfonamide | c1c(cc(cc1Cl)Cl)S(=O)(=O)N | 0103512963 |
| 2-aminoquinazolin-4-ol | c1ccc2c(c1)c(=O)[nH]c(n2)N | 0103512951 |
| 6-methyl-1,2,3,4-tetrahydropyridazin-3-one | CC1=NNC(=O)CC1 | 0103512952 |
| 6-methyl-2-phenyl-1,2,3,4-tetrahydropyridazin-3-one | CC1=NN(C(=O)CC1)C2=CC=CC=C2 | 0103512953 |
| 2-methylquinolin-6-amine | Cc1ccc2cc(ccc2n1)N | 0103512954 |
| 2-chlorobenzamide | NC(=O)c1ccccc1Cl | 0103512955 |
| 1-methyl-3-(trifluoromethyl)-4,5-dihydro-1H-pyrazol-5-one | CN1C(=O)CC(=N1)C(F)(F)F | 0103512956 |
| thiophene-2-carboxamide | NC(=O)c1cccs1 | 0103512957 |
| 3,5-dimethylpyrazin-2-amine | Cc1cnc(c(n1)C)N | 0103512958 |
| 2-(2,4-difluorophenoxy)pyridin-3-amine | c1cc(c(nc1)Oc2ccc(cc2F)F)N | 0103512959 |

| | | |
|---|---|---|
| 3-(benzyloxy)pyridin-2-amine | c1ccc(cc1)COc2cccnc2N | 0103512960 |
| 4-(4-chlorophenyl)piperidin-4-ol | c1cc(ccc1C2(CC[NH2+]CC2)O)Cl | 0103512102 |
| 4-methyl-3-phenyl-2,5-dihydro-1H-pyrrol-2-one | CC1=C(C(=O)NC1)c2ccccc2 | 0103512962 |
| methyl isoquinoline-3-carboxylate | COC(=O)c1cc2ccccc2cn1 | 0103512950 |
| pyridine-2,3-diamine | Nc1ncccc1N | 0103512949 |
| isoquinolin-3-amine | c1ccc2c[nH+]c(cc2c1)N | 0103512948 |
| phenyl(pyridin-2-yl)methanamine dihydrochloride | c1ccc(cc1)[C@H](c2ccccn2)[NH3+] | 0103512947 |
| 5-phenyl-1H-pyrazol-3-ol | C1C(=NNC1=O)C2=CC=CC=C2 | 0103512946 |
| 5-methyl-4H,7H-[1,2,4]triazolo[1,5-a]pyrimidin-7-one | Cc1cc(n2c(n1)ncn2)[O-] | 0103512945 |
| 6-phenylpyridazin-3-ol | c1ccc(cc1)c2ccc(=O)[nH]n2 | 0103512944 |
| quinolin-2-amine | c1ccc2c(c1)ccc(n2)N | 0103512943 |
| 1-phenyl-1,2,3,4-tetrahydroisoquinolin-3-one | c1ccc(cc1)[C@@H]2c3ccccc3CC(=O)N2 | 0103512942 |
| 1-[2-(1H-imidazol-5-yl)ethyl]pyrrolidine-2,5-dione | c1c([nH]c[nH+]1)CCN2C(=O)CCC2=O | 0103512941 |
| 2-methyl-4H-pyrido[1,2-a]pyrimidin-4-one | Cc1cc(=O)n2ccccc2n1 | 0103512940 |
| methyl 2-(5-methyl-2-phenyl-1,3-thiazol-4-yl)acetate | CC1=C(N=C(S1)C2=CC=CC=C2)CC(=O)OC | 0103512939 |
| 3-phenyl-1,2,4-thiadiazol-5-amine | Nc1nc(ns1)-c1ccccc1 | 0103512927 |
| 4-(2-methoxyphenyl)piperidine | COc1ccccc1C2CC[NH2+]CC2 | 0103512928 |
| 3-methyl-1-phenyl-1H-pyrazol-5-amine | Cc1cc(n(n1)c2ccccc2)N | 0103512929 |
| 2-(1H-indol-3-yl)ethan-1-ol | c1ccc2c(c1)c(c[nH]2)CCO | 0103512930 |
| 2,3-dihydro-1H-indole-2-carboxylic acid | c1ccc2c(c1)C[CH](N2)C(=O)[O-] | 0103512931 |
| (5-chloro-1-benzothiophen-3-yl)methanol | c1cc2c(cc1Cl)c(cs2)CO | 0103512932 |
| (5-methyl-3-phenyl-1,2-oxazol-4-yl)methanol | Cc1c(c(no1)c2ccccc2)CO | 0103512933 |
| [5-(pyridin-2-yl)thiophen-2-yl]methanol | OCC1=CC=C(S1)C2=CC=CC=N2 | 0103512934 |
| [5-(pyridin-2-yl)thiophen-2-yl]methanamine dihydrochloride | Cl.Cl.NCc1ccc(s1)-c1ccccn1 | 0103512935 |

| | | |
|---|---|---|
| pyrazin-2-ylmethanol | c1cnc(cn1)CO | 0103512936 |
| quinolin-6-ylmethanol | c1cc2cc(ccc2nc1)CO | 0103512937 |
| quinoline-4-carboxylic acid | c1ccc2c(c1)c(ccn2)C(=O)[O-] | 0103512938 |
| thieno[2,3-b]pyridin-2-ylmethanol | c1cc2cc(sc2nc1)CO | 0103512926 |
| methyl(thieno[2,3-b]pyridin-2-ylmethyl)ck)amine | CNCc1cc2cccnc2s1 | 0103512925 |
| (dimethyl-1,3-thiazol-5-yl)methanol | Cc1c(sc(n1)C)CO | 0103512924 |
| 1,3-benzothiazol-6-amine | c1cc2c(cc1N)scn2 | 0103512923 |
| 1,3-benzothiazol-2-ylmethanol | c1ccc2c(c1)nc(s2)CO | 0103512922 |
| 1,3-benzothiazol-2-ylmethanamine hydrochloride | c1ccc2c(c1)nc(s2)C[NH3+] | 0103512921 |
| methyl[(1-methyl-1H-indazol-3-yl)methyl]amine | CNCC1=N[N](C)C2=CC=CC=C12 | 0103512919 |
| 1-benzofuran-2-ylmethanol | OCc1cc2c(o1)cccc2 | 0103512918 |
| 1-benzofuran-5-ylmethanol | c1cc2c(cco2)cc1CO | 0103512917 |
| 1-benzofuran-5-amine | NC1=CC2=C(OC=C2)C=C1 | 0103512916 |
| 1,3-thiazol-2-ylmethanol | c1csc(n1)CO | 0103512915 |
| (5-phenyl-1,3-oxazol-4-yl)methanol | c1ccc(cc1)c2c(nco2)CO | 0103512903 |
| quinoxalin-6-amine | c1cc2c(cc1N)nccn2 | 0103512904 |
| (4-benzylmorpholin-2-yl)methanol | c1ccc(cc1)CN2CCO[CH](C2)CO | 0103512905 |
| (4-benzylmorpholin-2-yl)methanamine | c1ccc(cc1)CN2CCO[CH](C2)C[NH3+] | 0103512906 |
| (1,5-dimethyl-1H-pyrazol-3-yl)methanol | Cc1cc(nn1C)CO | 0103512907 |
| 1H-indole-3-carboxylic acid | c1ccc2c(c1)c(c[nH]2)C(=O)[O-] | 0103512908 |
| 1H-indole-2-carboxylic acid | c1ccc2c(c1)cc([nH]2)C(=O)[O-] | 0103512909 |
| [2-(pyridin-3-yl)-1,3-thiazol-4-yl]methanol | c1ccc2c(c1)cc([nH]2)C(=O)[O-] | 0103512910 |
| (5-methyl-1-phenyl-1H-pyrazol-4-yl)methanol | Cc1c(cnn1c2ccccc2)CO | 0103512911 |

| | | |
|---|---|---|
| [4-(morpholin-4-yl)phenyl]methanol | c1cc(ccc1CO)N2CCOCC2 | 0103512912 |
| 2-phenyl-1,3-thiazole-4-carboxylic acid | c1ccc(cc1)c2nc(cs2)C(=O)[O-] | 0103512913 |
| (2-phenyl-1,3-thiazol-4-yl)methanol | c1ccc(cc1)c2nc(cs2)CO | 0103512914 |
| (2-phenyl-1,3-thiazol-4-yl)methanamine | c1ccc(cc1)c2nc(cs2)C[NH3+] | 0103512105 |
| [4-(1H-pyrazol-1-yl)phenyl]methanol | OCc1ccc(cc1)n1cccn1 | 0103512106 |
| [4-(1H-imidazol-1-yl)phenyl]methanol | OCc1ccc(cc1)n1cncc1 | 0103512708 |
| [3-(2-methyl-1,3-thiazol-4-yl)phenyl]methanol | OCc1cccc(c1)c1csc(n1)C | 0103512707 |
| methyl([3-(2-methyl-1,3-thiazol-4-yl)phenyl]methyl)amine | CNCc1cccc(c1)c1csc(n1)C | 0103512706 |
| 6-phenoxypyridin-3-amine | Nc1ccc(nc1)Oc1ccccc1 | 0103512705 |
| 4-methyl-2-phenylpyrimidine-5-carboxylic acid | CC1=NC(=NC=C1C(O)=O)C2=CC=CC=C2 | 0103512704 |
| (4-methyl-2-phenylpyrimidin-5-yl)methanol | CC1=NC(=NC=C1CO)C2=CC=CC=C2 | 0103512703 |
| (3-methyl-5-phenyl-1,2-oxazol-4-yl)methanol | OCc1c(C)noc1c1ccccc1 | 0103512702 |
| (1-methyl-1H-imidazol-2-yl)methanol | OCc1nccn1C | 0103512701 |
| 6-phenylpyridine-3-carboxylic acid | OC(=O)C1=CC=C(N=C1)C2=CC=CC=C2 | 0103512700 |
| (6-phenylpyridin-3-yl)methanol | OCc1ccc(nc1)-c1ccccc1 | 0103512117 |
| 6-phenylpyridine-3-carbonitrile | N#Cc1ccc(nc1)-c1ccccc1 | 0103512687 |
| (2-methylimidazo[1,2-a]pyridin-3-yl)methanol | Cc1nc2ccccn2c1CO | 0103512688 |
| [3-(1H-pyrrol-1-yl)thiophen-2-yl]methanol | OCc1sccc1-n1cccc1 | 0103512689 |
| (1,3-dimethyl-1H-thieno[2,3-c]pyrazol-5-yl)methanol | Cc1nn(C)c2sc(CO)cc12 | 0103512690 |
| [5-(thiophen-2-yl)-1,2-oxazol-3-yl]methanol | OCc1cc(on1)-c1cccs1 | 0103512691 |
| methyl[(1-methyl-1H-imidazol-5-yl)methyl]amine | CNCc1cncn1C | 0103512692 |
| (1-benzyl-1H-imidazol-2-yl)methanol | OCc1nccn1Cc1ccccc1 | 0103512693 |
| 2,3-dihydro-1-benzofuran-2-ylmethanol | OCC1Cc2ccccc2O1 | 0103512694 |
| 2,3-dihydro-1-benzofuran-2-ylmethanamine hydrochloride | Cl.NCC1Cc2ccccc2O1 | 0103512695 |

| | | |
|---|---|---|
| 5-phenylpyridine-3-carboxylic acid | OC(=O)c1cncc(c1)-c1ccccc1 | 0103512696 |
| [2-(morpholin-4-yl)phenyl]methanol | OCc1ccccc1N1CCOCC1 | 0103512697 |
| 2-(1H-pyrazol-1-yl)benzonitrile | N#Cc1ccccc1-n1cccn1 | 0103512698 |
| 1-phenyl-1H-pyrazole-5-carboxylic acid | OC(=O)c1ccnn1-c1ccccc1 | 0103512686 |
| [4-(1H-pyrrol-1-yl)phenyl]methanol | OCc1ccc(cc1)-n1cccc1 | 0103512685 |
| [4-(1H-pyrrol-1-yl)phenyl]methanamine | NCc1ccc(cc1)-n1cccc1 | 0103512091 |
| 3,4-dihydro-2H-1-benzopyran-3-ylmethanol | OCC1COc2ccccc2C1 | 0103512683 |
| (dimethyl-1,3-oxazol-4-yl)methanol | Cc1nc(CO)c(C)o1 | 0103512682 |
| (1-methyl-5-phenyl-1H-pyrazol-4-yl)methanol | Cn1ncc(CO)c1-c1ccccc1 | 0103512681 |
| (2-methylquinolin-6-yl)methanol | Cc1ccc2cc(CO)ccc2n1 | 0103512680 |
| 1H-indol-4-ylmethanol | OCc1cccc2[nH]ccc12 | 0103512679 |
| methyl[(4-methyl-2-phenyl-1,3-thiazol-5-yl)methyl]amine | CNCc1sc(nc1C)-c1ccccc1 | 0103512678 |
| 3,4-dihydro-2H-1,4-benzoxazin-2-ylmethanol | OCC1CNc2ccccc2O1 | 0103512677 |
| (5-phenyl-1,2-oxazol-3-yl)methanol | OCc1cc(on1)-c1ccccc1 | 0103512676 |
| 4-methylthiophene-2-carboxamide | Cc1csc(c1)C(N)=O | 0103512675 |
| imidazo[1,2-a]pyridin-2-ylmethanol | OCc1cn2ccccc2n1 | 0103512663 |
| 3-(2-methylpyrimidin-4-yl)aniline | Cc1nccc(n1)-c1cccc(N)c1 | 0103512664 |
| imidazo[2,1-b][1,3]thiazol-6-ylmethanol | OCc1cn2ccsc2n1 | 0103512665 |
| 3-(1H-imidazol-1-yl)benzonitrile | N#Cc1cccc(c1)-n1ccnc1 | 0103512666 |
| 2-(1H-imidazol-1-yl)aniline | Nc1ccccc1-n1ccnc1 | 0103512667 |
| imidazo[1,2-a]pyridin-6-ylmethanol | OCc1ccc2nccn2c1 | 0103512668 |
| (imidazo[1,2-a]pyridin-6-ylmethyl)(methyl)amine | CNCc1ccc2nccn2c1 | 0103512669 |
| [3-(morpholin-4-yl)phenyl]methanol | OCc1cccc(c1)N1CCOCC1 | 0103512670 |
| methyl([3-(piperidin-1-yl)phenyl]methyl)amine | CNCc1cccc(c1)N1CCCCC1 | 0103512671 |

| | | |
|---|---|---|
| [2-(1H-pyrrol-1-yl)phenyl]methanol | OCc1ccccc1-n1cccc1 | 0103512672 |
| [4-(pyridin-4-yl)phenyl]methanol | OCc1ccc(cc1)-c1ccncc1 | 0103512673 |
| [3-(pyridin-4-yl)phenyl]methanol | OCc1cccc(c1)-c1ccncc1 | 0103512674 |
| 3-(pyridin-4-yl)aniline | Nc1cccc(c1)-c1ccncc1 | 0103512662 |
| methyl([3-(pyridin-4-yl)phenyl]methyl)amine | CNCc1cccc(c1)-c1ccncc1 | 0103512661 |
| [2-(4-methylpiperazin-1-yl)phenyl]methanol | CN1CCN(CC1)c1ccccc1CO | 0103512660 |
| 2-(4-methylpiperazin-1-yl)aniline | CN1CCN(CC1)c1ccccc1N | 0103512659 |
| methyl([3-(1,3-thiazol-2-yl)phenyl]methyl)amine | CNCc1cccc(c1)-c1nccs1 | 0103512658 |
| 3-(piperidin-1-yl)pyridine-2-carbonitrile | N#Cc1ncccc1N1CCCCC1 | 0103512657 |
| [3-(pyrrolidin-1-yl)phenyl]methanol | OCc1cccc(c1)N1CCCC1 | 0103512656 |
| methyl([3-(pyridin-3-yl)phenyl]methyl)amine | CNCc1cccc(c1)-c1cccnc1 | 0103512655 |
| [4-(pyridin-3-yl)phenyl]methanol | OCc1ccc(cc1)-c1cccnc1 | 0103512654 |
| [3-(morpholin-4-ylmethyl)phenyl]methanol | OCc1cccc(CN2CCOCC2)c1 | 0103512653 |
| methyl([3-(morpholin-4-ylmethyl)phenyl]methyl)amine | CNCc1cccc(CN2CCOCC2)c1 | 0103512652 |
| methyl([4-(1,3-thiazol-2-yl)phenyl]methyl)amine | CNCc1ccc(cc1)-c1nccs1 | 0103512651 |
| [4-(2-methyl-1,3-thiazol-4-yl)phenyl]methanol | Cc1nc(cs1)-c1ccc(CO)cc1 | 0103512639 |
| 2-(1-benzofuran-3-yl)ethan-1-ol | OCCc1coc2ccccc12 | 0103512640 |
| methyl(4-methyl-4H-thieno[3,2-b]pyrrol-5-ylmethyl)amine hydrochloride | Cl.CNCc1cc2sccc2n1C | 0103512641 |
| [4-(pyridin-2-yl)phenyl]methanol | OCc1ccc(cc1)-c1ccccn1 | 0103512642 |
| methyl([4-(pyridin-2-yl)phenyl]methyl)amine | CNCc1ccc(cc1)-c1ccccn1 | 0103512643 |
| [3-(pyrimidin-5-yl)phenyl]methanol | OCc1cccc(c1)-c1cncnc1 | 0103512644 |
| (2-methyl-4-phenyl-1,3-thiazol-5-yl)methanol | Cc1nc(c(CO)s1)-c1ccccc1 | 0103512645 |
| [3-(1H-imidazol-1-ylmethyl)phenyl]methanol | OCc1cccc(Cn2ccnc2)c1 | 0103512646 |
| [4-(1H-imidazol-1-ylmethyl)phenyl]methanol | OCc1ccc(Cn2ccnc2)cc1 | 0103512647 |

| | | |
|---|---|---|
| [2-(pyrrolidin-1-yl)pyridin-3-yl]methanamine | NCc1cccnc1N1CCCC1 | 0103512648 |
| [5-(thiophen-2-yl)-1H-pyrazol-3-yl]methanol | OCc1cc([nH]n1)-c1cccs1 | 0103512649 |
| (1-methyl-1H-indol-6-yl)methanol | Cn1ccc2ccc(CO)cc12 | 0103512650 |
| [4-(pyrimidin-5-yl)phenyl]methanol | OCc1ccc(cc1)-c1cncnc1 | 0103512638 |
| (2-methyl-2H-indazol-3-yl)methanol | Cn1nc2ccccc2c1CO | 0103512637 |
| methyl([2-(pyrrolidin-1-yl)phenyl]methyl)amine | CNCc1ccccc1N1CCCC1 | 0103512636 |
| 2-(piperazin-1-yl)aniline | Nc1ccccc1N1CCNCC1 | 0103512635 |
| [4-(piperidin-1-ylmethyl)phenyl]methanol | OCc1ccc(CN2CCCCC2)cc1 | 0103512634 |
| (4-[(4-methylpiperazin-1-yl)methyl]phenyl)methanol | CN1CCN(Cc2ccc(CO)cc2)CC1 | 0103512633 |
| 1H-1,3-benzodiazol-5-ylmethanol | OCc1ccc2[nH]cnc2c1 | 0103512632 |
| (1-methyl-5-phenyl-1H-pyrazol-3-yl)methanol | Cn1nc(CO)cc1-c1ccccc1 | 0103512631 |
| (1-methyl-5-phenyl-1H-pyrazol-3-yl)methanamine | Cn1nc(CN)cc1-c1ccccc1 | 0103512630 |
| (1-methyl-3-phenyl-1H-pyrazol-5-yl)methanol | Cn1nc(cc1CO)-c1ccccc1 | 0103512629 |
| methyl[(1-methyl-3-phenyl-1H-pyrazol-5-yl)methyl]amine | CNCc1cc(nn1C)-c1ccccc1 | 0103512628 |
| [1-methyl-5-(thiophen-2-yl)-1H-pyrazol-3-yl]methanol | Cn1nc(CO)cc1-c1cccs1 | 0103512627 |
| [1-methyl-3-(thiophen-2-yl)-1H-pyrazol-5-yl]methanol | Cn1nc(cc1CO)-c1cccs1 | 0103512090 |
| (2-[(4-methylpiperazin-1-yl)methyl]phenyl)methanol | CN1CCN(Cc2ccccc2CO)CC1 | 0103512616 |
| (3,5-dimethyl-1-phenyl-1H-pyrazol-4-yl)methanol | Cc1nn(c(C)c1CO)-c1ccccc1 | 0103512617 |
| [2-(morpholin-4-ylmethyl)phenyl]methanol | OCc1ccccc1CN1CCOCC1 | 0103512618 |
| 2-(morpholin-4-ylmethyl)aniline | Nc1ccccc1CN1CCOCC1 | 0103512619 |
| 2-(morpholin-4-ylmethyl)benzonitrile | N#Cc1ccccc1CN1CCOCC1 | 0103512620 |
| [4-(pyridin-2-yloxy)phenyl]methanol | OCc1ccc(Oc2ccccn2)cc1 | 0103512621 |
| 4-(pyridin-2-yloxy)benzonitrile | N#Cc1ccc(Oc2ccccn2)cc1 | 0103512622 |
| 5-methyl-2-phenylfuran-3-carboxylic acid | Cc1cc(C(O)=O)c(o1)-c1ccccc1 | 0103512623 |

| | | |
|---|---|---|
| (2-phenoxyphenyl)methanol | OCc1ccccc1Oc1ccccc1 | 0103512624 |
| methyl[(2-phenoxyphenyl)methyl]amine | CNCc1ccccc1Oc1ccccc1 | 0103512625 |
| [3-(pyridin-2-yloxy)phenyl]methanol | OCc1cccc(Oc2ccccn2)c1 | 0103512626 |
| [2-(piperidin-1-yl)pyridin-4-yl]methanol | OCc1ccnc(c1)N1CCCCC1 | 0103512118 |
| methyl([2-(piperidin-1-yl)pyridin-4-yl]methyl)amine | C[NH2+]Cc1ccnc(c1)N2CCCCC2 | 0103512901 |
| [2-(pyrrolidin-1-yl)pyridin-4-yl]methanol | c1c[nH+]c(cc1CO)N2CCCC2 | 0103512900 |
| (2-[2-(dimethylamino)ethoxy]phenyl)methanol | CN(C)CCOc1ccccc1CO | 0103512899 |
| (4-[2-(dimethylamino)ethoxy]phenyl)methanol | C[NH+](C)CCOc1ccc(cc1)CO | 0103512898 |
| (4-phenoxyphenyl)methanol | c1ccc(cc1)Oc2ccc(cc2)CO | 0103512897 |
| methyl[(4-phenoxyphenyl)methyl]amine | C[NH2+]Cc1ccc(cc1)Oc2ccccc2 | 0103512896 |
| 2-methyl-1,3-benzothiazol-6-amine | Cc1nc2ccc(cc2s1)N | 0103512895 |
| [6-(thiophen-2-yl)pyridin-3-yl]methanol | OCc1ccc(nc1)-c1cccs1 | 0103512894 |
| [4-(pyrrolidin-1-ylmethyl)phenyl]methanol | OCc1ccc(CN2CCCC2)cc1 | 0103512893 |
| 3-(pyrrolidin-1-ylmethyl)aniline | c1cc(cc(c1)N)C[NH+]2CCCC2 | 0103512892 |
| [4-(1H-pyrazol-1-ylmethyl)phenyl]methanol | OCc1ccc(Cn2cccn2)cc1 | 0103512124 |
| methyl([4-(1H-pyrazol-1-ylmethyl)phenyl]methyl)amine | CNCc1ccc(Cn2cccn2)cc1 | 0103512879 |
| methyl[(3-phenoxyphenyl)methyl]amine | C[NH2+]Cc1cccc(c1)Oc2ccccc2 | 0103512880 |
| methyl([2-(morpholin-4-yl)-5-(trifluoromethyl)phenyl]methyl)amine | CNCc1cc(ccc1N1CCOCC1)C(F)(F)F | 0103512881 |
| [2-(phenoxymethyl)phenyl]methanamine | c1ccc(cc1)OCc2ccccc2C[NH3+] | 0103512882 |
| 2H,3H,4H-pyrido[3,2-b][1,4]oxazine | C1COc2cccnc2N1 | 0103512883 |
| 3-(1H-pyrrol-1-yl)pyridine | c1ccn(c1)-c1cccnc1 | 0103512884 |
| 5-methyl-3-phenyl-1,2-oxazole-4-carboxamide | Cc1c(c(no1)c2ccccc2)C(=O)N | 0103512885 |
| 3-(1H-pyrrol-1-ylmethyl)pyridine | C(c1cccnc1)n1cccc1 | 0103512886 |
| 2-(1H-pyrrol-1-ylmethyl)pyridine | C(c1ccccn1)n1cccc1 | 0103512887 |

| | | |
|---|---|---|
| 4-phenyl-1,2,3-thiadiazole-5-carboxamide | c1ccc(cc1)c2c(snn2)C(=O)N | 0103512888 |
| 3-methylpyridin-2-ol | Cc1ccc[nH]c1=O | 0103512889 |
| 4-[(4-fluorophenyl)carbonyl]piperidine hydrochloride | c1cc(ccc1C(=O)C2CC[NH2+]CC2)F | 0103512890 |
| 2,4-difluorobenzene-1-sulfonamide | NS(=O)(=O)c1ccc(F)cc1F | 0103512878 |
| 4-(dimethyl-1,2-oxazole-4-sulfonyl)morpholine | Cc1c(c(on1)C)S(=O)(=O)N2CCOCC2 | 0103512877 |
| 3,5-dimethyl-4-(pyrrolidine-1-sulfonyl)-1,2-oxazole | Cc1noc(C)c1S(=O)(=O)N1CCCC1 | 0103512122 |
| 2-methyl-3-(1H-pyrazol-5-yl)pyrazine | Cc1c(nccn1)c2ccn[nH]2 | 0103512875 |
| 4-[(2-methyl-1,3-thiazol-4-yl)methoxy]benzonitrile | Cc1nc(COc2ccc(cc2)C#N)cs1 | 0103512874 |
| 4-benzyl-1,4-thiomorpholine-1,1-dione | c1ccc(cc1)CN2CCS(=O)(=O)CC2 | 0103512873 |
| N-[4-(propan-2-yl)phenyl]acetamide | CC(C)c1ccc(cc1)NC(=O)C | 0103512872 |
| 1-methyl-1H-pyrazol-3-amine | Cn1ccc(n1)N | 0103512871 |
| 5-methylthieno[2,3-d]pyrimidin-4-ol | Cc1csc2c1c(=O)[nH]cn2 | 0103512870 |
| 1-methyl-1H-pyrazol-5-amine | Cn1c(ccn1)N | 0103512869 |
| 3-(trifluoromethyl)-1H-pyrazole | FC(F)(F)c1cc[nH]n1 | 0103512868 |
| 2-[5-(trifluoromethyl)pyridin-2-yl]oxyacetonitrile | c1cc(ncc1C(F)(F)F)OCC#N | 0103512867 |
| 4-fluoro-1H-indazol-3-ol | Fc1cccc2[nH][nH]c(=O)c12 | 0103512855 |
| 4-methyl-N-[2-(2-oxoimidazolidin-1-yl)ethyl]benzamide | Cc1ccc(cc1)C(=O)NCCN2CCNC2=O | 0103512856 |
| 1-benzylpiperidin-4-ol | c1ccc(cc1)C[NH+]2CCC(CC2)O | 0103512857 |
| 1-benzyl-1H-imidazol-2-ol | O=c1[nH]ccn1Cc1ccccc1 | 0103512858 |
| 3-phenyl-1H-1,2,4-triazol-5-ol | c1ccc(cc1)c2[nH]c(=O)[nH]n2 | 0103512859 |
| 5-phenyl-1H-pyrazol-3-amine | c1ccc(cc1)c2cc(n[nH]2)N | 0103512860 |
| 2,5-dimethyl-4H,7H-pyrazolo[1,5-a]pyrimidin-7-one | Cc1cc2[nH]c(C)cc(=O)n2n1 | 0103512861 |
| 5-(thiophen-2-yl)-1H-pyrazol-3-amine | c1cc(sc1)c2cc([nH]n2)N | 0103512862 |
| 3-methyl-1,2-oxazol-5-amine | Cc1cc(N)on1 | 0103512863 |

| | | |
|---|---|---|
| 1-methanesulfonyl-4-methylbenzene | Cc1ccc(cc1)S(=O)(=O)C | 0103512864 |
| 1-methyl-3-phenyl-1H-pyrazol-5-amine | Cn1c(cc(n1)c2ccccc2)N | 0103512865 |
| 3-(4-chlorophenyl)-1H-pyrazol-5-amine | c1cc(ccc1c2cc(n[nH]2)N)Cl | 0103512866 |
| 4-(benzenesulfonyl)thiophen-3-amine | Nc1cscc1S(=O)(=O)c1ccccc1 | 0103512854 |
| 3-(4-bromophenyl)-1-methyl-1H-pyrazole | Cn1ccc(n1)c2ccc(cc2)Br | 0103512853 |
| 5-(pyridin-2-yl)thiophene-2-carboxamide | c1ccnc(c1)c2ccc(s2)C(=O)N | 0103513202 |
| 1-(2,3-dihydro-1H-indol-1-yl)ethan-1-one | CC(=O)N1CCc2c1cccc2 | 0103512851 |
| (1,4-dimethylpiperazin-2-yl)methanol | C[NH+]1CCN(C[C@H]1CO)C | 0103512850 |
| benzyl 3-oxopiperazine-1-carboxylate | c1ccc(cc1)COC(=O)N2CCNC(=O)C2 | 0103512849 |
| 1-benzyl-1H-imidazole | c1ccc(cc1)Cn2ccnc2 | 0103512848 |
| piperidin-4-ylmethanol | OCC1CCNCC1 | 0103512847 |
| 2-(methyl[2-(trifluoromethyl)quinolin-4-yl]amino)ethan-1-ol | CN(CCO)c1cc(nc2ccccc12)C(F)(F)F | 0103512846 |
| 6-methyl-4-(piperazin-1-yl)-2-(trifluoromethyl)quinoline | Cc1ccc2c(c1)c(cc([nH+]2)C(F)(F)F)N3CC[NH2+]CC3 | 0103512845 |
| 5-methyl-1,2-benzothiazole-3-carboxamide | CC1=CC2=C(C=C1)SN=C2C(=O)N | 0103512844 |
| (1-benzylpiperidin-4-yl)methanol | c1ccc(cc1)C[NH+]2CCC(CC2)CO | 0103512843 |
| 2-(pyridin-3-yl)ethan-1-ol | OCCc1cccnc1 | 0103512831 |
| 2-(4-benzylpiperazin-1-yl)ethan-1-amine | c1ccc(cc1)C[NH+]2CCN(CC2)CC[NH3+] | 0103512832 |
| 7-chloro-4-(piperazin-1-yl)quinoline | c1cc2c(ccnc2cc1Cl)N3CC[NH2+]CC3 | 0103512833 |
| 2,2-dimethyl-N-(pyridin-4-yl)propanamide | CC(C)(C)C(=O)Nc1ccncc1 | 0103512834 |
| 1,2-benzoxazol-3-amine | c1ccc2c(c1)c(no2)N | 0103512835 |
| 3,5-dimethyl-1-phenyl-1H-pyrazole-4-carboxylic acid | Cc1c(c(n(n1)c2ccccc2)C)C(=O)[O-] | 0103512836 |
| [1-(4-fluorophenyl)-5-methyl-1H-pyrazol-4-yl]methanol | Cc1c(cnn1c2ccc(cc2)F)CO | 0103512837 |
| 2,3-dihydro-1H-indol-2-ylmethanol | OCC1Cc2ccccc2N1 | 0103512838 |
| [2-(piperidin-1-yl)pyridin-3-yl]methanol | c1cc(c([nH+]c1)N2CCCCC2)CO | 0103512839 |

| | | |
|---|---|---|
| 3-[(pyrrolidin-1-yl)carbonyl]piperidine | C1CCN(C1)C(=O)[C@@H]2CCC[NH2+]C2 | 0103512840 |
| 4-[(pyrrolidin-1-yl)carbonyl]piperidine | C1CCN(C1)C(=O)C2CC[NH2+]CC2 | 0103512841 |
| [2-(3-chlorophenyl)-1,3-thiazol-4-yl]methanamine hydrochloride | Cl.NCc1csc(n1)-c1cccc(Cl)c1 | 0103512842 |
| 5-phenyl-1,2,4-thiadiazol-3-amine | Nc1nsc(n1)-c1ccccc1 | 0103512830 |
| [6-(piperidin-1-yl)pyridin-3-yl]methanol | c1cc([nH+]cc1CO)N2CCCCC2 | 0103512829 |
| 3,4-dihydro-2H-1-benzopyran-2-ylmethanol | c1ccc2c(c1)CC[C@H](O2)CO | 0103512828 |
| (6-aminopyridin-3-yl)methanol | c1cc(ncc1CO)N | 0103512827 |
| 1-(dimethyl-1,3-thiazole-5-sulfonyl)piperidine | Cc1nc(C)c(s1)S(=O)(=O)N1CCCCC1 | 0103512826 |
| 1-([5-methyl-2-(trifluoromethyl)furan-3-yl]carbonyl)piperidine | Cc1cc(C(=O)N2CCCCC2)c(o1)C(F)(F)F | 0103512825 |
| 4-(furan-2-sulfonyl)morpholine | O=S(=O)(N1CCOCC1)c1ccco1 | 0103512824 |
| 1-[(furan-3-yl)carbonyl]-4-methylpiperazine | CN1CCN(CC1)C(=O)c1ccoc1 | 0103512823 |
| 4-[(1-methyl-1H-pyrrol-2-yl)carbonyl]morpholine | Cn1cccc1C(=O)N1CCOCC1 | 0103512822 |
| 1-(1-methyl-1H-imidazole-4-sulfonyl)piperidine | Cn1cnc(c1)S(=O)(=O)N1CCCCC1 | 0103512821 |
| 5-[(pyrrolidin-1-yl)carbonyl]-1,2-oxazole | O=C(N1CCCC1)c1ccno1 | 0103512820 |
| N,N-dimethylquinoxaline-6-carboxamide | CN(C)C(=O)c1ccc2nccnc2c1 | 0103512819 |
| 4-[(1,3-dimethyl-1H-pyrazol-5-yl)carbonyl]morpholine | Cc1cc(C(=O)N2CCOCC2)n(C)n1 | 0103512807 |
| N,1,5-trimethyl-1H-pyrazole-3-carboxamide | CNC(=O)c1cc(C)n(C)n1 | 0103512808 |
| 4-(thiophene-2-sulfonyl)morpholine | O=S(=O)(N1CCOCC1)c1cccs1 | 0103512809 |
| N,N-dimethylthiophene-3-sulfonamide | CN(C)S(=O)(=O)c1ccsc1 | 0103512810 |
| 2,2-dimethyl-2,3-dihydro-1-benzofuran-7-carboxamide | CC1(C)Cc2cccc(C(N)=O)c2O1 | 0103512811 |
| N,N-dimethyloxane-4-carboxamide | CN(C)C(=O)C1CCOCC1 | 0103512812 |
| 1-[(oxan-4-yl)carbonyl]piperidine | O=C(C1CCOCC1)N1CCCCC1 | 0103512813 |
| 1-methyl-1H-pyrazole-3-carboxamide | Cn1ccc(n1)C(N)=O | 0103512814 |
| N-methyl-2-(oxan-4-yloxy)benzamide | CNC(=O)c1ccccc1OC1CCOCC1 | 0103512815 |

| | | |
|---|---|---|
| 5-chloro-N,1-dimethyl-1H-pyrazole-4-carboxamide | CNC(=O)c1cnn(C)c1Cl | 0103512816 |
| N,N,1,2-tetramethyl-1H-imidazole-4-sulfonamide | CN(C)S(=O)(=O)c1cn(C)c(C)n1 | 0103512817 |
| 1-[(3-fluorobenzene)sulfonyl]pyrrolidine | Fc1cccc(c1)S(=O)(=O)N1CCCC1 | 0103512818 |
| 2,2,6,6-tetramethyl-N-(2-methylphenyl)piperidin-4-amine | Cc1ccccc1NC2CC([NH2+]C(C2)(C)C)(C)C | 0103512096 |
| 5-acetyl-2,3,4,5-tetrahydro-1H-1,5-benzodiazepin-2-one | CC(=O)N1CCC(=O)Nc2c1cccc2 | 0103512094 |
| 1-(4-phenylpiperazin-1-yl)ethan-1-one | CC(=O)N1CCN(CC1)c2ccccc2 | 0103513284 |
| 2-phenyl-2,3,4,5-tetrahydro-1,4-benzoxazepin-5-one | O=C1NCC(Oc2ccccc12)c1ccccc1 | 0103513283 |
| 3-(2-methyl-1,3-thiazol-4-yl)pyridine | Cc1nc(cs1)-c1cccnc1 | 0103513282 |
| 3-phenyl-2,5-dihydro-1,2-oxazol-5-one | O=C1CC(=NO1)c1ccccc1 \|c:3\| | 0103513281 |
| 1,2-benzoxazol-3-ol | Oc1noc2ccccc12 | 0103513280 |
| 6-chloro-1,2-benzoxazol-3-ol | c1cc2c(cc1Cl)onc2[O-] | 0103513279 |
| 4-phenyl-1,3-thiazol-2-ol | c1ccc(cc1)c2csc(=O)[nH]2 | 0103513278 |
| 4-(trifluoromethyl)-1H,2H-pyrrolo[1,2-d][1,2,4]triazin-1-one | Oc1nnc(n2cccc12)C(F)(F)F | 0103512095 |
| 5-methyl-N-(pyridin-3-ylmethyl)-1,2-oxazol-3-amine | Cc1cc(NCc2cccnc2)no1 | 0103512099 |
| 4-phenyl-3-(pyridin-3-yl)-4,5-dihydro-1,2,4-oxadiazol-5-one | c1ccc(cc1)n2c(noc2=O)c3cccnc3 | 0103512101 |
| 6-phenyl-1,2,3,4-tetrahydropyridazin-3-one | c1ccc(cc1)C2=NNC(=O)CC2 | 0103513263 |
| 3-phenyl-[1,2,4]triazolo[4,3-b]pyridazin-6-amine | Nc1ccc2nnc(-c3ccccc3)n2n1 | 0103513264 |
| 6-methylidene-4-[(4-methylphenyl)methyl]-1,2,3,6-tetrahydropyridazin-3-one | Cc1ccc(cc1)Cc2cc(n[nH]c2=O)C | 0103513265 |
| 1-[4-(1H-indol-3-yl)piperidin-1-yl]ethan-1-one | CC(=O)N1CCC(CC1)c2c[nH]c3c2cccc3 | 0103513266 |
| 2-methyl-4-(piperazin-1-yl)quinoline | Cc1cc(c2ccccc2[nH+]1)N3CC[NH2+]CC3 | 0103513267 |
| 5-phenyl-1,3,4-thiadiazol-2-amine | c1ccc(cc1)c2nnc(s2)N | 0103513268 |
| 5-bromopyrimidin-2-amine | c1c(cnc(n1)N)Br | 0103513269 |
| 6-fluoro-1,3-benzothiazol-2-amine | c1cc2c(cc1F)sc(n2)N | 0103513270 |
| 5-phenyl-2,3-dihydro-1H-1,4-benzodiazepin-2-one | c1ccc(cc1)C2=NCC(=O)Nc3c2cccc3 | 0103513271 |

| | | |
|---|---|---|
| 3-methylcinnolin-5-amine | Cc1cc2c(cccc2nn1)N | 0103513272 |
| 3-methyl-5-phenylpyridazine | Cc1cc(cnn1)-c1ccccc1 | 0103513273 |
| 2,3-dihydro-1,4-benzodioxine-6-sulfonamide | NS(=O)(=O)c1ccc2OCCOc2c1 | 0103513274 |
| (6-methylpyridin-2-yl)methanol | Cc1cccc(n1)CO | 0103513262 |
| 2-methyl-6-phenylpyrimidin-4-ol | Cc1[nH]c(=O)cc(n1)c2ccccc2 | 0103513261 |
| 6-methyl-2-phenylpyrimidin-4-ol | Cc1cc(=O)[nH]c(n1)c2ccccc2 | 0103513260 |
| phthalazine-1,4-diol | c1ccc2c(c1)c(=O)[nH]nc2[O-] | 0103513259 |
| 2-(2-fluorophenyl)-1,3-thiazolidin-4-one | Fc1ccccc1C1NC(=O)CS1 | 0103513258 |
| [(4-fluorophenyl)methyl]urea | c1cc(ccc1CNC(=O)N)F | 0103513257 |
| 1H-indole | c1cc2ccccc2[nH]1 | 0103513256 |
| 5-methyl-1,2-oxazol-3-amine | Cc1cc(N)no1 | 0103513255 |
| dimethyl-1,2-oxazol-5-amine | Cc1c(noc1N)C | 0103513254 |
| 5-amino-2-phenyl-2,3-dihydro-1H-pyrazol-3-one | c1ccc(cc1)N2C(=O)CC(=N2)N | 0103513253 |
| 1-benzylpiperidin-4-amine | c1ccc(cc1)CN2CCC(CC2)[NH3+] | 0103513252 |
| 1H-1,3-benzodiazol-2-ol | c1ccc2c(c1)[nH]c(=O)[nH]2 | 0103513251 |
| 5-methylpyridin-2-amine | Cc1ccc([nH+]c1)N | 0103513239 |
| pyridin-4-amine | c1c[nH+]ccc1N | 0103513240 |
| 5-methyl-1,3,4-thiadiazol-2-amine | Cc1nnc(N)s1 | 0103513241 |
| 1,3-dimethylimidazolidin-2-one | CN1CCN(C1=O)C | 0103513242 |
| 2-methylbenzene-1-sulfonamide | Cc1ccccc1S(=O)(=O)N | 0103513243 |
| 2-(4-hydroxyphenyl)acetamide | c1cc(ccc1CC(=O)N)O | 0103513244 |
| 2-phenyl-1H-imidazole-4-carboxylic acid | c1ccc(cc1)c2[nH]cc(n2)C(=O)[O-] | 0103513245 |
| 5-(4-methylphenyl)-1,3-oxazole | Cc1ccc(cc1)c2cnco2 | 0103513246 |
| 4-(1,3-oxazol-5-yl)pyridine | c1ncc(o1)-c1ccncc1 | 0103513247 |

| | | |
|---|---|---|
| 3-(1,3-oxazol-5-yl)aniline | c1cc(cc(c1)N)c2cnco2 | 0103513248 |
| 2-(4-methylphenyl)-1,3,4-oxadiazole | Cc1ccc(cc1)c2nnco2 | 0103513249 |
| 5-(4-methylphenyl)-1,2,4-thiadiazole | Cc1ccc(cc1)c2ncns2 | 0103513250 |
| 2-(1H-pyrazol-3-yl)pyridine | c1cc(n[nH]1)-c1ccccn1 | 0103513238 |
| 3,5-dimethyl-1-phenyl-1H-pyrazole-4-carboxamide | Cc1c(c(n(n1)c2ccccc2)C)C(=O)N | 0103513237 |
| 5-phenyl-1,3-oxazole | c1ccc(cc1)c2cnco2 | 0103513236 |
| 4-[(2,6-difluorobenzene)sulfonyl]morpholine | c1cc(c(c(c1)F)S(=O)(=O)N2CCOCC2)F | 0103513235 |
| 4-phenyl-1H-imidazole | c1ccc(cc1)c2cnc[nH]2 | 0103513234 |
| 5-fluoroquinazolin-4-amine | c1cc2c(c(c1)F)c(ncn2)N | 0103513233 |
| ethyl 2,5-dimethyl-1H-pyrrole-3-carboxylate | CCOC(=O)c1cc([nH]c1C)C | 0103513232 |
| N-(4-fluoro-1-methyl-1H-indazol-3-yl)propanamide | CCC(=O)Nc1nn(C)c2cccc(F)c12 | 0103513231 |
| 2-hydroxy-6-methylpyridine-3-carbonitrile | Cc1ccc(C#N)c(=O)[nH]1 | 0103513230 |
| 6-methoxypyridine-3-carbonitrile | COc1ccc(cn1)C#N | 0103513229 |
| 3-benzyl-2H,3H-[1,3]oxazolo[4,5-b]pyridin-2-one | c1ccc(cc1)Cn2c3c(cccn3)oc2=O | 0103513228 |
| (2-methoxyphenyl)thiourea | COc1ccccc1NC(N)=S | 0103513227 |
| 2-methoxypyridin-3-amine | COc1c(cccn1)N | 0103513215 |
| N-methyl-[1,3]thiazolo[5,4-b]pyridin-2-amine | CNc1nc2cccnc2s1 | 0103513216 |
| 3,4-dimethylbenzamide | Cc1ccc(cc1C)C(=O)N | 0103513217 |
| methyl(pyridin-3-ylmethyl)amine | C[NH2+]Cc1cccnc1 | 0103513218 |
| N-(pyridin-3-ylmethyl)acetamide | CC(=O)NCc1cccnc1 | 0103513219 |
| 1-methyl-1H-imidazole-4-sulfonamide | Cn1cc(nc1)S(=O)(=O)N | 0103513220 |
| 4-phenylquinazolin-2-ol | c1ccc(cc1)c2c3ccccc3nc(=O)[nH]2 | 0103513221 |

**Appendix Figure 3b: Sigma Fragment Library Collection**

| CHEMICAL NAME | SMILES | BARCODE |
|---|---|---|
| pyridine-2-carboxamide | NC(=O)C1=NC=CC=C1 | 103559462 |
| [2-(3-chlorophenyl)-1,3-thiazol-4-yl]methanol | O=C1COc2cccnc2N1 | B00050367 |
| 5-(4-fluorophenyl)-1,2-oxazol-3-amine | Nc1cc(on1)-c1ccc(F)cc1 | 103559461 |
| 1-[(furan-2-yl)carbonyl]piperazine | O=C(N1CCNCC1)c1ccco1 | 103559460 |
| 4-phenyl-1,2,4-triazolidine-3,5-dione | O=c1[nH][nH]c(=O)n1-c1ccccc1 | 103559459 |
| 5-phenyl-1H-pyrazole-4-carboxylic acid | OC(=O)c1cn[nH]c1-c1ccccc1 | 103559458 |
| 2H,4H-pyrido[3,2-b][1,4]oxazin-3-one | O=C1COc2cccnc2N1 | 103559457 |
| isoquinolin-1-ol | C1=CC=C2C(=C1)C=CNC2=O | 103559456 |
| pyrimidine-2,4-diamine | C1=CN=C(N=C1N)N | 103559455 |
| 3-phenyl-1H-pyrazole-5-carboxylic acid | C1=CC=C(C=C1)C2=NNC(=C2)C(=O)O | 103559454 |
| 6-chloro-2-N-ethyl-1,3,5-triazine-2,4-diamine | CCNc1nc(N)nc(Cl)n1 | 103559453 |
| 5-chloro-1,3-benzoxazol-2-amine | C1=CC2=C(C=C1Cl)N=C(O2)N | 103559452 |
| 4-benzylpiperazin-2-one | C1CN(CC(=O)N1)CC2=CC=CC=C2 | 103559451 |
| 5-chloro-2-methyl-1,3-benzoxazole | Cc1nc2cc(Cl)ccc2o1 | 103559439 |
| quinolin-2-ol | C1=CC=C2C(=C1)C=CC(=O)N2 | 103559440 |
| 5,5-dimethyl-1,3-oxazolidine-2,4-dione | CC1(C)OC(=O)NC1=O | 103559441 |
| 6-methyl-9H-purine | CC1=C2C(=NC=N1)N=CN2 | 103559442 |
| N-methylpyridin-2-amine | CNc1ccccn1 | 103559443 |
| 1H-pyrazolo[3,4-d]pyrimidin-4-ol | C1=C2C(=NC=NC2=O)NN1 | 103559444 |
| 5-chloro-1H-1,3-benzodiazole | Clc1ccc2[nH]cnc2c1 | 103559445 |
| 1,3-benzoxazol-2-ol | Oc1nc2ccccc2o1 | 103559446 |
| 1H-indazole-3-carboxylic acid | C1=CC=C2C(=C1)C(=NN2)C(=O)O | 103559447 |

| | | |
|---|---|---|
| 5-bromopyrimidin-2-ol | C1=CC(=O)NC=C1Br | 103559448 |
| 3-phenyl-1,2-oxazol-5-amine | Nc1cc(no1)-c1ccccc1 | 103559449 |
| phthalazin-1-ol | Oc1nncc2ccccc12 | 103559450 |
| 1H-pyrazolo[3,4-d]pyrimidin-4-amine | C1=NNC2=C1C(=NC=N2)N | 103559438 |
| 6-phenyl-1,3,5-triazine-2,4-diamine | C1=CC=C(C=C1)C2=NC(=NC(=N2)N)N | 103559437 |
| 5-chloropyridin-2-ol | Oc1ccc(Cl)cn1 | 103559436 |
| 3-chloro-1H-indazole | ClC1=N[NH]C2=C1C=CC=C2 | 103559435 |
| N-benzylpyridin-2-amine | C1=CC=C(C=C1)CNC2=CC=CC=N2 | 103559434 |
| quinazolin-4-ol | C1=CC=C2C(=C1)C(=O)N=CN2 | 103559433 |
| quinoline-3-carboxylic acid | C1=CC=C2C(=C1)C=C(C=N2)C(=O)O | 103559432 |
| benzoic acid amine | N.OC(=O)c1ccccc1 | 103559431 |
| 4-methyl-2-phenyl-1H-imidazole | CC1=CN=C(N1)C2=CC=CC=C2 | 103559430 |
| 3,4-dihydro-2H-1,4-benzoxazin-3-one | C1C(=O)NC2=CC=CC=C2O1 | 103559429 |
| quinoxalin-2-ol | C1=CC=C2C(=C1)NC(=O)C=N2 | 103559428 |
| 1H-1,3-benzodiazole-5-carboxylic acid | [O-]C(=O)c1ccc2[nH]cnc2c1 | 103559427 |
| N-ethylpyridine-3-carboxamide | CCNC(=O)C1=CC=CN=C1 | 103559415 |
| 2-methylquinoline-4-carboxamide | CC1=NC2=CC=CC=C2C(=C1)C(=O)N | 103559416 |
| N-(2-methyl-1,3-benzoxazol-5-yl)acetamide | CC1=NC2=C(O1)C=CC(=C2)NC(=O)C | 103559417 |
| N,1-diethyl-2-methyl-1H-1,3-benzodiazole-5-sulfonamide | CCNS(=O)(=O)c1ccc2n(CC)c(C)nc2c1 | 103559418 |
| 5-methyl-3-phenyl-1H-pyrazole-4-carboxamide | CC1=C(C(=NN1)C2=CC=CC=C2)C(=O)N | 103559419 |
| 1-ethyl-5-methanesulfonyl-2-methyl-1H-1,3-benzodiazole | CCN1C(=NC2=C1C=CC(=C2)S(=O)(=O)C)C | 103559420 |
| 3-cyclohexyl-1-(pyridin-3-ylmethyl)urea | C1CCC(CC1)NC(=O)NCC2=CN=CC=C2 | 103559421 |
| N-(pyrimidin-2-yl)acetamide | CC(=O)NC1=NC=CC=N1 | 103559422 |
| 1-(4-fluorophenyl)-3-methyl-1H-pyrazol-5-ol | Cc1cc(O)n(n1)-c1ccc(F)cc1 | 103559423 |
| 5-(4-methylphenyl)-1,2-oxazole | CC1=CC=C(C=C1)C2=CC=NO2 | 103559424 |

| | | |
|---|---|---|
| 7,8-dimethyl-2,3,4,5-tetrahydro-1H-1,5-benzodiazepin-2-one | Cc1cc2NCCC(=O)Nc2cc1C | 103559425 |
| N-(pyridin-2-yl)acetamide | CC(=O)NC1=CC=CC=N1 | 103559426 |
| 6-aminopyridine-3-sulfonamide | C1=CC(=NC=C1S(=O)(=O)N)N | 103514126 |
| 6-bromo-2-methylpyrazolo[1,5-a]pyrimidine | CC1=NN2C=C(C=NC2=C1)Br | 103514066 |
| 2-hydroxy-1-propyl-1H-1,3-benzodiazole-5-sulfonamide | CCCn2c(O)nc1cc(S(N)(=O)=O)ccc12 | 103513586 |
| 6-bromo-[1,2,4]triazolo[1,5-a]pyrimidine | C1=C(C=NC2=NC=NN21)Br | 103513658 |
| 2-methyl-4-phenyl-1,3-thiazole-5-carboxylic acid | CC1=NC(=C(S1)C(=O)O)C2=CC=CC=C2 | 103513634 |
| 1,2-benzoxazol-3-ylmethanesulfonamide | C1=CC=C2C(=C1)C(=NO2)CS(=O)(=O)N | 103514114 |
| 5-(2-chlorophenyl)-1H-1,2,4-triazole | C1=CC=C(C(=C1)C2=NC=NN2)Cl | 103513611 |
| N-methyl-3,4-dihydro-2H-1,4-benzoxazine-2-carboxamide | CNC(=O)C1CNC2=CC=CC=C2O1 | 103513646 |
| N-(pyridin-2-ylmethyl)acetamide hydrochloride | CC(=O)NCC1=CC=CC=N1.Cl | 103514103 |
| 2-(benzyloxy)-6-chloropyrazine | C1=CC=C(C=C1)COC2=CN=CC(=N2)Cl | 103514102 |
| pyrido[2,3-d]pyrimidin-4-ol | C1=CC2=C(NC=NC2=O)N=C1 | 103514078 |
| 6-fluoro-1,2,3,4-tetrahydroquinolin-2-one | FC1=CC2=C(NC(=O)CC2)C=C1 | 103514150 |
| 2-chloro-5-phenyl-1,3,4-thiadiazole | C1=CC=C(C=C1)C2=NN=C(S2)Cl | 103512073 |
| 1-(propane-1-sulfonyl)pyrrolidine-2-carboxylic acid (note chiral) | CCCS(=O)(=O)N1CCC[C@H]1C(O)=O | 103513647 |
| 1-phenyl-1H-pyrazol-3-ol | C1=CC=C(C=C1)N2C=CC(=O)N2 | 103513623 |
| 1-(4-chlorophenyl)-1H-1,2,3,4-tetrazole | Clc1ccc(cc1)-n1cnnn1 | 103512060 |
| 1-benzyl-3-methyl-1H-pyrazol-5-amine | CC1=NN(C(=C1)N)CC2=CC=CC=C2 | 103514055 |
| 3,6-dimethyl-1-benzofuran-2-carboxylic acid | CC1=CC2=C(C=C1)C(=C(O2)C(=O)O)C | 103513610 |
| 2-amino-1,3-benzothiazole-6-sulfonamide | C1=CC2=C(C=C1S(=O)(=O)N)SC(=N2)N | 103513622 |
| 5-methyl-7-phenyl-4H,5H,6H,7H-[1,2,4]triazolo[1,5-a]pyrimidine | CC1CC(N2C(=N1)N=CN2)C3=CC=CC=C3 | 103514127 |
| 3-amino-1-6,2-benzothiazole-1,1-dione | NC1=N[S](=O)(=O)C2=C1C=CC=C2 | 103514090 |
| 3-ethyl-4-oxo-3,4-dihydrophthalazine-1-carboxylic acid | CCN1C(=O)C2=CC=CC=C2C(=N1)C(=O)O | 103513575 |
| 3,6-dimethyl-[1,2]oxazolo[5,4-b]pyridine-4-carboxylic acid | Cc1noc2nc(C)cc(C(O)=O)c12 | 103513598 |
| 1,3,6-trimethyl-1H-pyrazolo[3,4-b]pyridine-4-carboxylic acid | CC1=NC2=C(C(=NN2C)C)C(=C1)C(=O)O | 103514079 |

| | | |
|---|---|---|
| 1-[(4-fluorophenyl)methyl]-1H-pyrazol-4-amine | Nc1cnn(Cc2ccc(F)cc2)c1 | 103513599 |
| [5-(4-chlorophenyl)-1,2,4-oxadiazol-3-yl]methanamine | C1=CC(=CC=C1C2=NC(=NO2)CN)Cl | 103514067 |
| 1-(3-phenyl-2H,4H,5H,6H,7H-pyrazolo[4,3-c]pyridin-5-yl)propan-1-one | CCC(=O)N1CCc2n[nH]c(c2C1)-c1ccccc1 | 103514091 |
| 1-[(4-methylphenyl)methyl]-1H-1,2,4-triazol-3-amine | Cc1ccc(Cn2cnc(N)n2)cc1 | 103513587 |
| 2-(piperidin-1-ylmethylidene)-1H,2H,3H,4H-thieno[3,2-d]pyrimidin-4-one | O=C1NC(CN2CCCCC2)NC3=C1SC=C3 | 103514138 |
| 3-methyl-4-phenyl-1,2-oxazol-5-amine | CC1=NOC(=C1C2=CC=CC=C2)N | 50476863 |
| [2-(3-chlorophenyl)-1,3-thiazol-4-yl]methanol | OCc1csc(n1)c1cccc(c1)Cl | B00050367 |

**Appendix figures 3a and 3b.** A tabulated version of the entire 433-Maybridge-Sigma fragment screening library used in this study. Chemical vendors, chemical names, SMILES (Simplified Molecular Input Line Entry System – a notation used in organic chemistry to depict chemical structures and information) and barcodes have been provided for each individual fragment to enable ease of location.

**Appendix figure 4 –Software/programs used routinely throughout this study**

| Software/program | Use | Reference | Websource |
|---|---|---|---|
| **NMRPipe** | Processing NMR spectra | F. Delaglio et al. 1995 | https://spin.niddk.nih.gov/NMRPipe/ |
| **PDB** | Visualising protein crystal structures | H.M. Berman et al. 2000 | http://www.rcsb.org/pdb/home/home.do |
| **Sedfit** | Analysis of AUC data | P Schuck et al. 2000[199, 200] | http://analyticalultracentrifugation.com/default.htm |
| **Protparam** | Calculating protein molecular weight, protein concentrations and determination of protein pIs | E Gasteiger et al. 2005 | http://web.expasy.org/protparam/ |
| **CCPN** | Analysis of NMR spectra | WF Vranken et al. 2005 | http://www.ccpn.ac.uk/ |
| **VNMRJ** | Guide for use of NMR spectrometers and experimental selection | - | http://openvnmrj.org/ |
| **CDSSTR** | Analysis of CD data | Sreerama and Woody, 2000 | http://dichroweb.cryst.bbk.ac.uk/html/home.shtml |
| **Biacore** | Analysis of SPR data | - | https://www.biacore.com/lifesciences/index.html |